

Obrint dades a l'Arxiu Nacional de Catalunya

Enric Terradellas Prat

Director del treball: Remo Suppi Boldrito

Any: 2016

Treball d'Investigació del Màster en Gestió Documental, Transparència i Accés a la Informació

Escola Superior d'Arxivística i Gestió de Documents

Col·lecció: Treballs final de màster i de postgrau

Com citar aquest treball: Terradellas Prat, Enric. (2016) *Obrint dades a l'Arxiu Nacional de Catalunya*. Treball d'Investigació del Màster en Gestió Documental, Transparència i Accés a la Informació de l'Escola Superior d'Arxivística i Gestió de Documents de la Universitat Autònoma de Barcelona. (Treballs final de màster i de postgrau). Http://... (consultat el ...)



Esta obra està sujeta a licencia Creative Commons Reconocimiento-NoComercial-SinObrasDerivadas 3.0 España (<http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>).

Se permite la reproducción total o parcial i la comunicació pública de la obra, siempre que no sea con finalidades comerciales, y siempre que se reconozca la autoría de la obra original. No se permite la creación de obras derivadas.

Títol: Obrint dades a l'Arxiu Nacional de Catalunya

Resum

En els darrers temps han irromput amb força un seguit de conceptes com *open data*, participació ciutadana o reutilització de la informació que ja formen part del vocabulari habitual de la societat del segle XXI. La filosofia *open data* té per objectiu posar a disposició de la societat les dades que gestiona l'administració pública en formats fàcils de manipular per tal que qualsevol persona pugui analitzar, reutilitzar i redistribuir aquestes dades, generant nous serveis i permetent que l'administració pública millori en transparència i govern obert.

Els arxius públics, com a dipositants finals de la documentació generada per les administracions, no poden ser aliens a aquesta realitat i han de ser proactius en la promoció i difusió, "en obert", de les dades que gestionen i dels seus instruments de descripció.

El present treball té com a objectiu plantejar les claus teòriques que han de permetre als arxius facilitar al ciutadà informació que pugui ser reutilitzada, i fer un cas d'estudi per valorar la viabilitat d'implantar un projecte de dades obertes en una institució arxivística com l'Arxiu Nacional de Catalunya.

Paraules clau: dades obertes, participació ciutadana, reutilització de la informació, transparència, govern obert, arxius, Arxiu Nacional de Catalunya.

Título: Abriendo datos en el Arxiu Nacional de Catalunya

Resumen

En los últimos tiempos han irrumpido con fuerza una serie de conceptos como *open data*, participación ciudadana o reutilización de la información que ya forman parte del vocabulario habitual de la sociedad del siglo XXI. La filosofía *open data* tiene como objetivo poner a disposición de la sociedad los datos que gestiona la administración pública en formatos fáciles de manipular para que cualquier persona pueda analizar, reutilizar y redistribuir estos datos, generando nuevos servicios y permitiendo que la administración pública mejore en transparencia y gobierno abierto.

Los archivos públicos, como depositantes finales de la documentación generada por las administraciones, no pueden ser ajenos a esta realidad y deben ser proactivos en la promoción y difusión, "en abierto", de los datos que gestionan y de sus instrumentos de descripción.

El presente trabajo tiene como objetivo plantear las claves teóricas que deben permitir a los archivos facilitar al ciudadano información que pueda ser reutilizada, y elaborar un caso de estudio para valorar la viabilidad de implantar un proyecto de datos abiertos en una institución archivística como el Arxiu Nacional de Catalunya.

Palabras clave: datos abiertos, participación ciudadana, reutilización de la información, transparencia, gobierno abierto, archivos, Arxiu Nacional de Catalunya.

Title: Opening data in the National Archive of Catalonia

Abstract

Recently terms like open data, citizen participation or reuse of the information have burst strongly and they are already part of the usual vocabulary of the XXI Century society.

Open data philosophy aims to make available to society public data in easy formats to be manipulated, analyzed, reused and redistributed, generating new services and allowing public administration to get better in transparency and open government.

Public archives, as final depositors of the records generated by the administrations, can not ignore this reality, and must be proactive in the promotion and dissemination, in an open way, of their catalogues and data managed by them.

This paper aims to raise the theoretical keys that should allow archives providing information to citizens that can be reused, and elaborate a case study to assess the feasibility of implementing an open data project in an archivist institution like the National Archive of Catalonia.

Keywords: open data, citizen participation, reuse of information, transparency, open government, archives, National Archive of Catalonia.

SUMARI

1	Introducció.....	7
1.1	Context i justificació.....	7
1.2	Objectius	9
2	Marc teòric i estat de la qüestió	11
2.1	Dades obertes.....	11
2.1.1	El concepte de dades obertes	11
2.1.2	Formats més habituals de dades obertes.....	12
2.1.3	Com obrir dades.....	13
2.1.4	Per què obrir dades?.....	15
2.1.5	El concepte Linked Open Data: el 5 estrelles dels formats en obert	16
2.1.6	La Web Semàntica	19
2.2	Legislació relativa a la reutilització de la informació	20
2.2.1	Directives europees sobre reutilització de la informació del sector públic.....	20
2.2.2	Lleis estatals sobre reutilització de la informació del sector públic.....	21
2.2.3	Altres lleis de l'estat espanyol relacionades amb la RISP	22
2.3	Plataformes de publicació de dades obertes	23
2.3.1	Principals plataformes disponibles al mercat.....	24
2.4	Iniciatives governamentals d'obertura de dades.....	26
2.4.1	Portals de dades obertes estatals	28
2.4.2	El portal de dades obertes de la Generalitat de Catalunya	29
2.5	Experiències d'institucions arxivístiques en obertura de dades	30
2.5.1	Iniciatives internacionals	30
2.5.2	Iniciatives arreu de l'estat espanyol.....	36
2.5.3	El protocol OAI-PMH i l' <i>Open Archive Initiative</i>	41
3	L'Arxiu Nacional de Catalunya.....	43
3.1	Missió i funcions.....	44

3.2	Organització	45
3.3	Serveis que ofereix	45
3.4	Fons que s'hi custodien.....	46
3.5	Activitats i publicacions	48
3.6	La pàgina web de l'ANC	48
3.7	Els cercadors	50
3.7.1	CercANC i Arxius en línia	50
3.8	El sistema informàtic de l'Arxiu Nacional de Catalunya	52
3.8.1	GANC (Gestor de l'Arxiu Nacional de Catalunya).....	52
3.8.2	La fitxa de primers tractaments i el format d'intercanvi de dades	53
3.9	Algunes dades de l'Arxiu Nacional de Catalunya	54
3.9.1	Volum total de documentació custodiada a l'Arxiu Nacional de Catalunya	54
3.9.2	Volum total de documentació descrita a l'Arxiu Nacional de Catalunya	55
3.9.3	Volum total de documents consultables a Internet (Arxius en Línia)	56
3.9.4	Usuaris de l'ANC i consultes de les descripcions	56
4	Proposta d'obertura de dades a l'Arxiu Nacional de Catalunya	57
4.1	Com es mostra la informació a Arxius en Línia.....	58
4.1.1	Algunes mancances d'Arxius en Línia pels usuaris del segle XXI	63
4.2	Disposar de formats reutilitzables: una demanda a l'alça	64
4.3	Definició d'un projecte d'Open Data per l'Arxiu Nacional de Catalunya	66
4.3.1	Quines dades s'utilitzaran per fer la prova de concepte	67
4.3.2	La legislació permet obrir aquestes dades?	68
4.3.3	Termes d'ús i llicències de les dades que es proposen oferir	69
4.3.4	El formats en què es proposa oferir les dades	70
4.3.5	Plataforma sobre la que es farà la prova de concepte.....	70
4.3.6	Resultats de la prova de concepte realitzada sobre CKAN	71
4.3.7	Valoració de la prova de concepte i ús de CKAN per obrir dades a l'ANC.....	77
5	Conclusions.....	79
6	Bibliografia i fonts.....	80

1 Introducció

1.1 Context i justificació

Durant la segona dècada de l'actual segle XXI han irromput amb força una sèrie de conceptes i tendències marcats amb paraules com dades obertes (*open data*), dades massives (*big data*), transparència o participació ciutadana que han assenyalat l'inici d'una nova era en la gestió i utilització de la informació.

Les dades i la informació continguda en aquestes han passat a tenir un paper primordial en la societat actual i han facilitat l'apoderament del ciutadà. Si fins fa uns anys el tractament i ús de les dades quedava reduït en entorns de gestió i govern, l'obertura en la provisió i el flux de les mateixes n'ha permès la seva reutilització, generant nous continguts i serveis que han permès que el ciutadà pugui incrementar la seva capacitat de triar, actuar i construir una nova realitat social en la que és un actor actiu consumint i generant nova informació.

El moviment *open* i la cultura lliure han condicionat l'entorn digital de gestió i publicació de la informació tal i com se la coneixia fins ara i han precipitat l'adopció de diferents accions encaminades a ordenar i legislar tot allò relacionat amb la reutilització de la informació.

Així, la Unió Europea regula, des de l'any 2003 i mitjançant la Directiva 2003/98/CE (modificada per la Directiva 2013/37/UE), la reutilització de la informació del sector públic, mentre que l'estat espanyol ho fa des de l'any 2007, a partir de la Llei 37/2007 (modificada per la Llei 18/2015), que té en compte que la informació generada des de les instàncies públiques ofereix una enorme potencialitat i és de gran interès per a les empreses, que poden contribuir al creixement econòmic i a la creació de llocs de treball, i per als ciutadans, com a element de transparència i guia per a la participació democràtica.

Amb tot, la Reutilització de la Informació del Sector Públic (RISP) és l'objectiu principal de la iniciativa *Open Data*, que consisteix a posar la informació del sector públic disponible i en formats estàndard oberts, facilitant-ne el seu accés i permetent-ne la seva reutilització tant a empreses com a particulars. Val a dir que les dades obertes no només tenen perquè ser governamentals o d'institucions públiques, sinó que també ho poden ser d'empreses privades o fruit de la col·laboració entre les primeres i les segones.

Els arxius, com a ens de caràcter públic i dipositants finals de la documentació generada per les administracions públiques, no poden ser aliens a aquesta nova realitat i han d'adoptar

una actitud proactiva en la promoció i la difusió, “en obert”, dels seus instruments de descripció.

Les institucions arxivístiques disposen de dades correctament estructurades, de qualitat, certes, contextualitzades i autèntiques i poden oferir als ciutadans grans volums d'informació reutilitzable, és a dir, fitxers de dades que puguin ser descarregats pels ciutadans i que aquests en puguin explotar el contingut segons els seus interessos.

Si bé la presència a Internet dels arxius ha augmentat molt significativament en els darrers anys, posant a l'abast dels usuaris la descripció dels fons i documents que custodien i facilitant-ne l'accessibilitat, ara és el moment de fer un pas més, de guanyar visibilitat i de posar en valor la feina que es fa des dels arxius, adoptant un paper més actiu i incorporant el gran volum de dades que s'hi gestionen als portals de dades obertes que ja existeixen o “obrint” la informació que es facilita actualment a les pàgines web dels arxius per tal que pugui ser reutilitzada. En definitiva, fer que els instruments de descripció que elaboren els serveis d'arxiu esdevinguin dades obertes.

L'Arxiu Nacional de Catalunya (ANC) és la institució cabdal del país encarregada d'aplegar, conservar i difondre el patrimoni documental de Catalunya. Amb la voluntat d'afavorir la consolidació d'una societat lliure, oberta i democràtica assegura l'accés dels ciutadans als documents que s'hi custodien, garantint-ne els seus drets i el control per part dels mateixos de les actuacions dels poders públics.

Un dels serveis principals que ofereix l'ANC és el de permetre la consulta pública i gratuïta dels seus instruments de descripció i fons documentals, essent la seva pàgina web i l'aplicatiu Arxius en Línia la porta principal d'accés a les descripcions que l'Arxiu posa a disposició dels usuaris, havent quedat relegat a un segon terme el servei d'atenció més tradicional, és a dir, el de tipus presencial. Això es així perquè, com moltes institucions arxivístiques, l'ANC ha fet un gran esforç en els últims anys per permetre la recuperació de la informació en línia, posant a l'abast de qualsevol ciutadà les descripcions dels fons i documents conservats per la institució.

Malgrat això, i tal i com s'acaba de comentar, el context actual fa que l'ANC hagi d'anar més enllà, bo i oferint la possibilitat als seus usuaris de reutilitzar la informació continguda en els inventaris i catàlegs que estan disponibles en línia per tal de facilitar la feina als investigadors en les seves recerques o afavorint la creació de nous productes que puguin ser d'utilitat per la ciutadania en general.

1.2 Objectius

Aquest treball té la voluntat, d'una banda, de plantejar les claus teòriques que han de permetre als arxius aportar les seves descripcions a través de portals de dades obertes i RISP per tal que la seva informació pugui ser reutilitzada en qualsevol àmbit, ja sigui la recerca, la difusió o, fins i tot, l'explotació econòmica; i de l'altra, fer un cas d'estudi valorant la viabilitat d'implantar un projecte de dades obertes en una institució arxivística com l'Arxiu Nacional de Catalunya.

Per fer-ho, es plantejaran les possibilitats que tenen els arxius històrics de Catalunya, i més en concret l'Arxiu Nacional de Catalunya, institució que es considera capdavantera dels arxius de la Generalitat de Catalunya, per dur a terme iniciatives de dades obertes, ja sigui des del portal Arxius en Línia –cercador que permet accedir de forma conjunta i des d'un únic entorn web, als fons i documents que es custodien a l'Arxiu Nacional de Catalunya, als arxius comarcals i als històrics provincials–, ja sigui aportant conjunts de *datasets* al portal de dades obertes de què disposa la Generalitat de Catalunya, o ja sigui avaluant la possibilitat d'utilitzar plataformes *open source* de publicació de dades obertes.

Amb tot, el treball s'estructurarà en els següents apartats:

1. Marc teòric i estat de la qüestió: en aquest primer apartat es tractaran els següents temes:
 - El concepte de dades obertes: significat, importància i criteris que han de complir les dades per considerar-les obertes.
 - Formats més habituals dels arxius utilitzats en *open data*: s'enumeraran els formats més freqüents en dades obertes i s'explicaran les seves característiques principals.
 - Com obrir dades: quins són els criteris que s'han de seguir en el moment d'obrir dades segons el *World Wide Web Consortium* (W3C) i els autors del manual "*The Open Data Handbook*".
 - Per què obrir dades: raons per les quals una institució hauria d'obrir dades i exemples que mostren la potencialitat d'aquesta obertura.
 - Breu explicació del concepte Linked Open Data i la Web Semàntica.
 - Legislació relativa a la reutilització de la informació: en aquest punt es farà un anàlisi de la legislació referent a la reutilització de la informació del sector públic (RISP), tant a nivell europeu com a nivell estatal i es farà un breu anàlisi d'altres lleis de l'estat espanyol relacionades amb la RISP.
 - Plataformes que permeten la publicació de dades en obert: es detallaran les principals plataformes existents que permeten publicar dades en obert, peça

clau a l'hora de donar el suport tecnològic necessari en una estratègia de dades obertes per part de qualsevol administració o organisme públic.

- Iniciatives governamentals d'obertura de dades: s'explicarà com les diferents institucions governamentals estan desenvolupant portals específics per oferir les seves dades obertes al conjunt de la ciutadania a nivell mundial, europeu i estatal.
- Experiències d'institucions arxivístiques en obertura de dades: es mostraran casos reals d'institucions arxivístiques d'arreu del món i de l'estat espanyol que ja estan oferint els seus instruments de descripció en format obert per tal que pugui ser reutilitzada la informació continguda en ells per qui ho consideri oportú. D'altra banda, també s'exposaran quins altres tipus de dades aporten els arxius als portals de dades obertes dels seus països respectius.
- El protocol OAI-PMH i l'*Open Archive Initiative*: partint d'alguns dels exemples citats a l'apartat anterior es farà una explicació d'aquest protocol i d'aquesta iniciativa i les oportunitats que ofereixen als arxius.

2. L'Arxiu Nacional de Catalunya: en el segon gran apartat del treball es parlarà sobre:

- La història, missió, visió i valors de l'Arxiu Nacional de Catalunya.
- Com s'organitza l'ANC, quins serveis ofereix i els fons que s'hi custodien.
- Com és la seva pàgina web.
- Quines són les característiques principals del seu cercador.
- Quina ha estat l'evolució del sistema informàtic dels arxius històrics de la Generalitat de Catalunya i més en concret de l'ANC: el GANC, la fitxa de primers tractaments i el format d'intercanvi de dades.
- Quina ha estat l'activitat de l'Arxiu Nacional de Catalunya entre els anys 2011 i 2015 a partir de l'exposició de dades relatives a: volum de documentació custodiada, volum total de la documentació descrita, total de documents consultables a Internet i número d'usuaris i de consultes fetes de les descripcions.

3. Proposta d'obertura de dades a l'Arxiu Nacional de Catalunya: finalment es farà un plantejament de projecte d'obertura de dades per l'ANC. Per fer-ho:

- S'exposarà com es mostra actualment la informació a Arxius en Línia i s'enumeraran algunes de les seves mancances.
- Es justificarà, amb demandes reals que ha rebut l'Arxiu, l'oportunitat d'oferir la informació que hi ha a Arxius en Línia en formats oberts.
- Es plantejarà la viabilitat de dur a terme un projecte d'aquest tipus en una institució com l'Arxiu Nacional de Catalunya.

- Es mostraran els resultats d'una prova de concepte realitzada sobre la plataforma CKAN d'un conjunt d'arxius que descriuen fons custodiats per l'Arxiu Nacional de Catalunya.
- 4. El darrer apartat del treball estarà destinat a exposar les conclusions extretes a partir de l'elaboració del present treball.

2 Marc teòric i estat de la qüestió

2.1 Dades obertes

2.1.1 El concepte de dades obertes

Dades obertes o *open data* és un terme que es refereix a l'exposició pública d'informació de manera adequada per al seu accés i lliure reutilització per la ciutadania, empreses i altres organismes. És un moviment que parteix de la base que la informació de l'administració és patrimoni de tots els ciutadans i que aquesta informació els ha de ser retornada.

Dades obertes comprèn tant les dades que el sector públic exposa públicament, com el privat, essent més adequat parlar d'obertura d'informació del sector públic o *open government data* pel primer cas.

Aquest últim concepte es va definir d'una forma clara l'any 2007 (MALAMUD, 2007), després de celebrar-se una reunió d'experts a Estats Units on es discutia com el govern podia obrir les dades que tenia emmagatzemades electrònicament per al seu ús públic. D'aquesta reunió en va sorgir una llista de 8 principis bàsics que, d'aplicar-se, potenciaria l'ús públic de les dades del govern. Aquesta llista de 8 principis bàsics seria ampliada, l'any 2010, fins a 10 principis per la *Fundació Sunlight* (WONDERLICH, 2010). Aquests principis són els que s'enumeren tot seguit:

1. Integritat: el conjunt de dades ha de ser el més complet possible sobre un tema en particular, les dades han d'estar anonimitzades, juntament amb les seves metadades, fórmules i dades derivades, si les hi ha.
2. Dades primàries: informació original recollida pel govern i detalls de com es va recollir per tal que l'usuari sàpiga com es va fer i el grau de precisió amb el qual es va fer.
3. Disponibilitat: sempre que sigui possible i tant aviat com s'hagi recollit sempre prioritzant aquelles dades temporals sensibles a aquesta variable.
4. Facilitat d'accés: minimitzar totes les barreres que permetin accedir a les dades.
5. Accessibles per ordinadors: facilitar formats que puguin ser llegits pels ordinadors a més d'altres formats més entenedors pels humans.

6. No discriminació: eliminar les barreres d'ús de les dades (inscripcions, registres, etc.).
7. Ús d'estàndards oberts: les dades hauran de ser publicades en formats oberts per permetre que tothom hi tingui accés (si es publiquen en un format propietari se'n derivaran costos associats ja que s'haurà de tenir el programa propietari per obrir-los).
8. Llicències: reduir al màxim la imposició de "Termes de servei", restriccions de difusió o d'atribució, ja que aquestes restriccions actuaran com una barrera per a la lliure utilització de les dades. S'ha d'etiquetar, doncs, aquesta informació com a obra del govern i disponible sense restriccions d'ús com a part de domini públic.
9. Permanència: capacitat de trobar la informació al llarg del temps, per la qual cosa haurà d'estar en línia i a perpetuïtat. En cas que canviï, incloure una línia de temps o de versions.
10. Cost: s'han d'eliminar els costos per l'usuari que vol reutilitzar les dades publicades ja que la informació publicada ha estat recollida amb finalitats governamentals i el cost imposat als usuaris té molt poc o nul efecte sobre l'acció del govern en el moment de recollir les dades.

Amb tot, es pot afirmar que el concepte dades obertes fa referència a dades que poden ser utilitzades, reutilitzades i redistribuïdes lliurement per qualsevol individu, facilitant la interoperabilitat entre dades que permeti obtenir-ne de noves que possibilitin la generació de nous serveis i productes o simplement nova informació. Tot plegat amb la voluntat de facilitar la vida als ciutadans i permetre guiar-los en la presa de decisions.

Les dades obertes, en definitiva, han de respondre a les següents característiques:

- Disponibilitat i accessibilitat: el conjunt ha de ser sencer, sense costos associats i en formats oberts i modificables.
- Reutilitzables i redistribuïbles: s'ha de poder utilitzar, distribuir i combinar amb altres dades.
- Accés no discriminatori: qualsevol individu ha de tenir accés i possibilitat d'utilitzar, reutilitzar i redistribuir la informació sense etiquetes que en redueixin la seva possibilitat d'ús.

2.1.2 Formats més habituals de dades obertes

Una de les característiques més importants de les dades obertes és que els seus formats siguin oberts, és a dir, que no es necessiti cap eina propietària per poder-hi accedir. Aquests formats també han de ser llegibles per ordinadors i modificables.

Els formats més utilitzats en *open data* són:

- **JSON** (*JavaScript Object Notation*): format d'estàndard obert que utilitza text per identificar els objectes i els seus atributs com a parelles de valors. És fàcil d'entendre per a les persones i ofereix senzillesa a les màquines en la generació i interpretació. L'extensió de l'arxiu en format Json és json.
- **XML** (*Extensible Markup Language*): format similar al JSON, desenvolupat pel W3C i definit com a estàndard lliure obert. Es caracteritza per la seva simplicitat, generalitat i usabilitat a través d'Internet i és àmpliament utilitzat per a representar dades en estructures de qualsevol forma i habitualment utilitzades en serveis Web.
- **HTML** (*HyperText Markup Language*): format de dades obertes caracteritzat per la seva facilitat i cost reduït per ser generat. En l'actualitat hi ha molts conjunts de dades que utilitzen taules en html per especificar les dades o algun document addicional que expliqui com extreure les dades de determinada pàgina.
- **RDF** (*Resource Description Framework*): és un model estàndard per l'intercanvi de dades a la Web i facilita la fusió de dades. RDF amplia l'estructura d'enllaços web utilitzant URI's¹ per indicar la relació entre objectes, així com els dos extrems de les connexions. Aquest model permet que dades estructurades i semiestructurades puguin ser barrejades, exposades i compartides a través de diferents aplicacions. És un format complex, però és considerat un dels formats amb major futur en l'entorn de la web semàntica i l'*open data*.
- **CSV** (*Comma-Separated Values*): format que s'utilitza per representar dades en forma de taula, en les que les columnes es separen per comes i les files per salts de línia. Normalment s'ha d'acompanyar de documentació addicional que expliqui què representa cada columna o fila per tal que les dades siguin útils.
- **Office Open XML**: format d'arxiu obert i estàndard les extensions més comunes del qual són .docx, .xlsx i .pptx. S'utilitza per representar i emmagatzemar fulles de càlcul, gràfiques, presentacions i documents de text. Un arxiu d'aquest tipus conté, principalment, dades basades en el llenguatge de marcat XML i comprimides en un contenidor .zip específic.

2.1.3 Com obrir dades

L'any 2009 el *World Wide Web Consortium* (W3C), consorci internacional dirigit per Tim Berners-Lee i que treballa per a desenvolupar i promocionar estàndards per al *World Wide Web*, va publicar una guia de bones pràctiques per ajudar a les institucions públiques d'arreu del món a alliberar les seves dades (BENNET, D. i HARVEY, A., 2009) i on es descriuen les passes a seguir per a la seva publicació:

¹ Un URI (*Uniform Resource Identifier*) és una cadena de caràcters que identifica els recursos d'una xarxa de manera unívoca. La diferencia respecte una URL (localitzador de recursos uniforme) és que aquests últims fan referència a recursos que, generalment, poden variar en el temps.

1. Publicar les dades en brut, mantenint una estructura que en permeti el seu ús automatitzat.
2. Crear un catàleg en línia per tal que els usuaris hi puguin accedir i sàpiguen què s'ha publicat. Les dades que s'hi publiquen han de ser fiables, estructurades i documentades.
3. Enriquir les dades en format xhtml (*Extensible HyperText Markup Language*), format que combina la sintaxis del html, dissenyat per mostrar les dades, i l'xml, dissenyat per descriure'ls. La finalitat de l'ús de l'xhtml és el d'avançar en el desenvolupament de la web semàntica, on la informació i la forma de presentar-la es troben separades.

D'altra banda, segons els autors del manual "*The Open Data Handbook*", existeixen 3 regles fonamentals per transformar les dades en obertes:

1. Simplicitat: començar amb un conjunt de dades petit, senzill i ràpid. Es pot tractar d'un conjunt de dades o una part d'un gran conjunt de dades.
2. Rapidesa: publicar les dades tan aviat com sigui possible. Això generarà confiança, compromís, fiabilitat i expectació.
3. Gestionar dubtes i temors: aclarir els dubtes que puguin sorgir dins la mateixa institució o organització el més ràpidament possible, per evitar malentesos i desconfiança.

A la vegada, i segons els mateixos autors del manual acabat de citar, existeixen 4 passes principals per a la creació de dades obertes:

1. Tria del conjunt de dades: per fer-ho es poden tenir en compte una sèrie de consideracions:
 - Preguntar-se si és factible publicar determinades dades.
 - Preguntar a la comunitat (potencials usuaris d'aquestes dades) per obtenir una idea de la demanda que hi pugui haver.
 - Avaluar el valor que poden tenir les dades a obrir.
 - Mirar què estan fent i quines dades estan obrint altres institucions similars.
2. Ús d'una llicència oberta: utilitzar algunes de les llicències alineades amb la "definició d'obert".
3. Posar el conjunt de dades a disposició de tothom qui hi estigui interessat: en un format obert i llegible per ordinador per facilitar-ne la seva reutilització. Així, es pot oferir la informació en format pdf però s'ha de tenir en compte que si es facilita en aquest format la informació serà fàcilment comprensible pels humans però difícilment reutilitzable per un ordinador. En aquest punt també cal valorar si les dades es penjen a la mateixa pàgina web de què disposa la institució, si s'utilitzen portals on

és possible publicar-los de manera gratuïta o si es desenvolupen aplicacions (API's) per mostrar determinada informació.

4. Fer detectable el *dataset* publicat: assegurar que la gent pugui trobar la font del material. Per fer-ho, existeixen eines a Internet dissenyades específicament per fer més detectables les dades. Una de les més prominents és [Datahub](#), un catàleg i magatzem de bases de dades de tot el món que facilita a les organitzacions publicar materials i als usuaris trobar-los.

2.1.4 Per què obrir dades?

Les dades obertes, i en especial les dades obertes dels governs o institucions públiques, són un gran recurs degut a la qualitat i gran quantitat d'informació que contenen. Les administracions públiques, en l'exercici de les seves funcions, recullen gran quantitat de dades que contenen informació molt valuosa que, combinada amb altra informació i reutilitzada, pot aportar valor i facilitar la vida dels ciutadans, obtenint aquests últims un retorn de la informació que puguin haver cedit a les administracions.

A més, molta de la informació governamental és pública per llei (a excepció d'aquella que conté dades de caràcter personal especialment protegides i aquella no accessible perquè salvaguarda interessos públics) i, en conseqüència, pot ser oberta i posada a disposició dels ciutadans.

En el sector públic, tenir accés a les dades de l'administració garanteix la transparència, l'eficiència i la igualtat d'oportunitats a l'hora de crear valor. Transparència perquè es poden consultar i tractar dades que venen directament de les fonts oficials; eficiència perquè ciutadans i organitzacions poden crear serveis de forma més ajustada en col·laboració amb l'administració; i igualtat d'oportunitats perquè l'accés és el mateix per a tothom.

Actualment ja es poden enumerar una llarga llista d'àrees on l'obertura de dades governamentals estan creant valor: transparència i control democràtic, participació, autoapoderament, millora i creació de nous productes i serveis, innovació, millora en l'eficiència dels serveis oferts pels governs o mesurament de l'impacte de les polítiques, entre d'altres.

Existeixen exemples en la majoria d'aquestes àrees:

- Transparència: el projecte del govern britànic "[Where does my Money go?](#)" mostra com el govern està gastant els diners recaptats a partir dels impostos dels contribuents.
- Ajuda a la presa de decisions personals: serveis com "[Mapumental](#)" al Regne Unit permeten saber quin és el millor lloc per viure en funció de la durada del viatge fins al lloc de treball.

- Millora en l'eficiència governamental: el ministeri d'educació holandès va publicar a Internet totes les dades relacionades amb educació per a la seva reutilització. Des d'aquell moment, el número de consultes que rep el ministeri relacionades amb l'educació ha baixat, reduint el volum de treball i els costos per l'administració.

Amb tot, es pot afirmar que l'obertura de dades per part de les administracions públiques té una potencialitat enorme, transformant, amb petites inversions, les dades de què disposen les institucions públiques en informació útil pel ciutadà i la pròpia administració, i fomentant la generació de nous models de negoci basats en la informació, que es traduiran en riquesa i nous llocs de treball.

Obrir dades proporciona transparència en el seu estat més pur, perquè en les dades hi ha l'origen de la presa de decisions i, per tant, són bàsics per contribuir a la transparència i per explicar els perquès de les decisions (DE LOS RÍOS, N., 2013).

2.1.5 El concepte Linked Open Data: el 5 estrelles dels formats en obert

L'any 2009 Tim Berners-Lee, l'inventor de la Web, explica el concepte de dades enllaçades en una famosa conferència TED (*Technology, Entertainment and Design*), animant a particulars i a organitzacions a obrir les seves dades a la web.

Un any més tard, el mateix Berners-Lee, proposa una escala de 5 estrelles, de menor a major obertura de les dades:

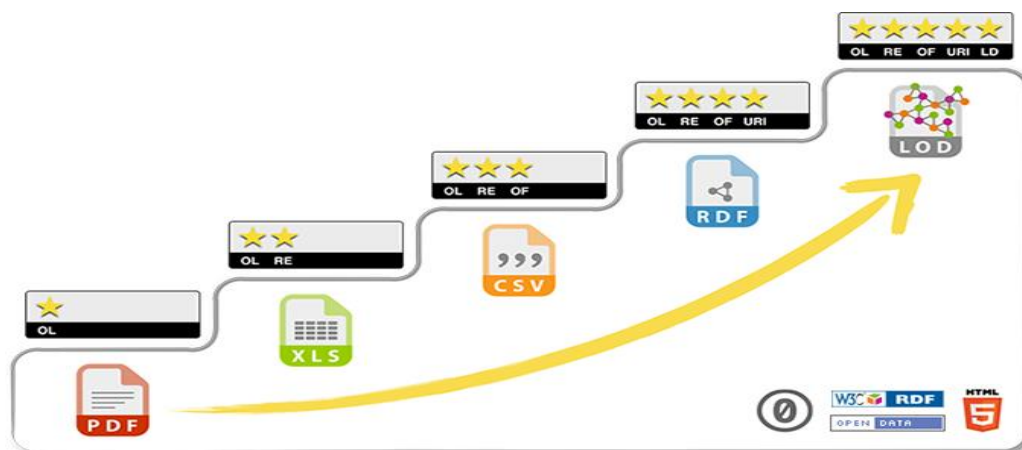


Figura 1. Esquema de desenvolupament de 5 estrelles per les dades obertes. Font: 5startdata.info

Estrelles	Significat	Avantatges	Inconvenients	Exemple de fitxers
1	Publicar dades a la web en qualsevol format i sota una llicència oberta	Les dades són accessibles a la web sota una llicència oberta	Les dades estan atrapades en un document i es requereix la creació d'un software per extreure-les	pdf
2	Publicar dades estructurades	Les dades són accessibles a la Web de forma estructurada	Les dades estan atrapades en un document i es necessita un software propietari per extreure-les	xls
3	Utilitzar formats no propietaris per publicar les dades	Les dades són accessibles a la Web i qualsevol persona els pot utilitzar fàcilment	Encara només són dades publicades a la Web i no són dades integrades a la Web	csv
4	Utilitzar URI's (que és una direcció Web d'una dada que serveix per enllaçar-la amb altres dades) per identificar coses i propietats, de manera que es pugui apuntar a aquestes dades	Les dades estan integrades a la Web, tenen una URI i poden ser compartides a la Web		rdf
5	Enllaçar les teves dades a altres dades que hi ha a la Web per dotar-les de context. És a dir, que la informació publicada a la Web apunti a altres fonts de dades	Les dades estan integrades a la Web i enllaçades amb altres dades.		lod

Taula 1. Proposta de classificació de les dades obertes segons Tim Berners-Lee. Font: 5startdata.info

El concepte *Linked Data* o dades enllaçades descriu un mètode de publicació de dades estructurades que puguin ser interconnectades i més útils. Els quatre principis que caracteritzen les dades enllaçades són:

1. Utilitzar URI's per identificar els recursos publicats a la Web.
2. Aprofitar l'HTTP de l'URI per tal que qualsevol persona pugui localitzar i consultar els recursos publicats a la Web.
3. Proporcionar informació útil sobre els recursos publicats a la Web quan l'URI hagi estat localitzat i consultat.
4. Incloure enllaços a altres URI's relacionats amb les dades contingudes en el recurs, de manera que es potenciï el descobriment d'informació a la Web.

Les dades enllaçades es representen a través de models RDF (*Resource Description Framework*), format explicat en el punt 2.1.2 d'aquest mateix treball. Aquest és un model de dades genèric, basat en tripletes (subjecte-predicat-objecte) també anomenats grafs (objecte-propietat-valor), que permet estructurar i enllaçar objectes o coses en el món i expressar informació sobre recursos (ja siguin documents, objectes físics, persones, etc.). El subjecte representaria un recurs o entitat, el predicat designaria una propietat d'aquest subjecte, i l'objecte expressaria el valor d'aquesta propietat.

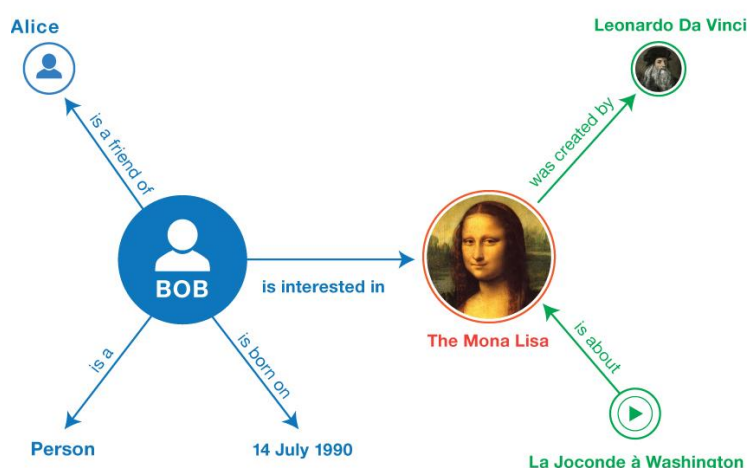


Figura 2. Exemple de graf basat en tripletes. Font: [W3C](http://www.w3.org)

És un format pensat perquè les aplicacions puguin processar informació sobre la Web i aquesta última pugui ser intercanviada sense pèrdua de significat. És, en definitiva, l'estàndard que s'utilitza per descriure els recursos web, especificar metadades i representar informació.

El seu llenguatge es recolza en una base sintàctica, normalment XML, metallenguatge extensible d'etiquetes. La seva validació a través de *XML Schema* permet ser descodificat pels ordinadors.

Un arxiu *open data* en format RDF és considerat *Linked Open Data* i és l'estàndard W3C per la web semàntica, un projecte iniciat pel mateix Tim Berners-Lee per la creació de tecnologies per publicar dades llegibles per aplicacions informàtiques unint metadades semàntiques i ontològiques a Internet (BERNERS-LEE, T., HENDLER, J., LASSILA, O., 2001). Amb la web semàntica les administracions poden enllaçar dades provinents de diverses fonts, institucions o organitzacions, explorar i combinar aquestes dades lliurement i sense restriccions.

2.1.6 La Web Semàntica

La web semàntica o web de dades és el futur de la Web segons Tim Berners-Lee. Segons aquesta visió, a la Web s'hi publicaran múltiples bases de dades distribuïdes, obertes i enllaçades entre elles. La web semàntica permetrà augmentar la interoperabilitat entre sistemes, possibilitant l'automatització de moltes de les tasques que actualment desenvolupen humans. D'altra banda, millorarà els sistemes de cerca i recuperació a la Web, identificant els recursos d'informació unívocament i utilitzant vocabularis que ajudin a descriure el recurs i el seu context, proporcionant una semàntica apta per a que es pugui processar de manera automàtica.

Per fer-ho possible cal desenvolupar mecanismes que garanteixin una interoperabilitat completa basada en tres criteris: interoperabilitat tècnica (protocols), sintàctica (estructura) i semàntica (contingut). Els tres nivells d'interoperabilitat es relacionen en una dependència jeràrquica.

Els principals components de la web semàntica són els metallenguatges i estàndards de representació XML, *XML Schema*, RDF, *RDF Schema* i OWL, així com el llenguatge SPARQL per a la consulta de dades RDF:

- XML: aporta la sintaxi superficial pels documents estructurats.
- *XML Schema*: és un llenguatge per definir l'estructura dels documents XML.
- RDF: aporta infraestructura de dades i vocabularis. Com s'ha dit anteriorment, és un model de dades pels recursos i les relacions que es poden establir entre ells.
- *RDF Schema*: vocabulari per descriure les propietats i les classes dels recursos RDF, amb una semàntica per establir jerarquies de generalització entre propietats i classes.
- OWL: llenguatge d'ontologies per al web.
- SPARQL: és un llenguatge de consulta de conjunts de dades RDF. És l'estàndard de consulta per a la web semàntica i les regles d'inferència (*rules*) o conjunt de condicions basades sobre axiomes definits a les ontologies que permeten a les aplicacions semàntiques com raonadors o agents inferir coneixement.

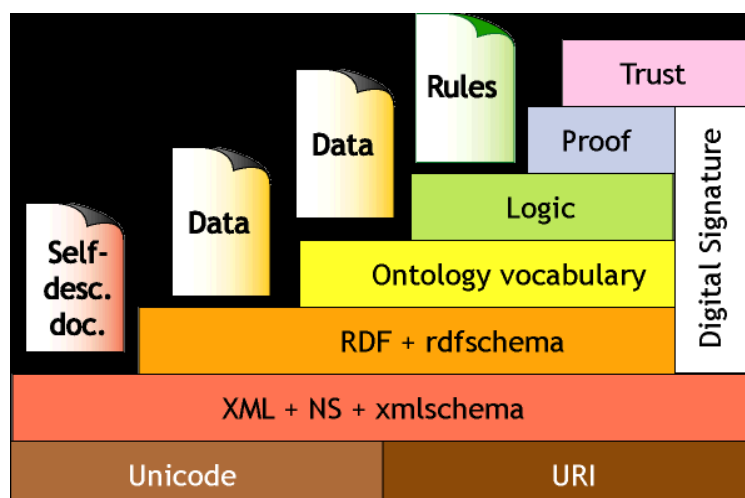


Figura 3. Arquitectura tecnològica de la Web Semàntica. Font: [W3C](#)

2.2 Legislació relativa a la reutilització de la informació

Tant a nivell europeu com a nivell de l'estat espanyol existeixen directives i normes que regulen les principals qüestions relacionades amb les dades obertes. Totes elles estan orientades a promoure l'obertura de dades per part de les administracions i a garantir l'accés dels ciutadans a la informació pública i la seva reutilització.

2.2.1 Directives europees sobre reutilització de la informació del sector públic

La primera referència legislativa relacionada amb la reutilització de la informació del sector públic a nivell europeu la trobem l'any 2003, moment en què apareix la Directiva 2003/98/CE², que estableix un conjunt mínim de normes que regulen la reutilització i els instruments pràctics que facilitin la reutilització dels documents existents conservats per organismes del sector públic dels Estats membres.

Passats deu anys de l'aparició de la primera directiva europea relativa a la reutilització de la informació i degut al fet que durant aquest temps transcorregut s'incrementa exponencialment el volum de dades a tot el món i apareixen noves tecnologies que permeten analitzar i explotar aquestes dades, es considera oportú promulgar una nova directiva que modifiqui la de l'any 2003. Així és com, l'any 2013, apareix publicada la Directiva 2013/37/UE³.

Aquesta directiva té per objectiu facilitar la creació de productes i serveis d'informació basats en la reutilització de dades del sector públic, garantint l'ús, en diferents països de la UE, de

² [Directiva 2003/98/CE](#) del Parlament Europeu i del Consell, de 17 de novembre de 2003, relativa a la reutilització de la informació del sector públic.

³ [Directiva 2013/37/UE](#) del Parlament Europeu i del Consell, de 26 de juny de 2013, per la que es modifica la Directiva 2003/98/CE relativa a la reutilització de la informació del sector públic.

dades del sector públic a empreses i ciutadans i promoure la lliure circulació de la informació i la comunicació, garantint la protecció de les dades personals, la propietat intel·lectual i industrial i dotant a tot el conjunt de la seguretat jurídica necessària per a la reutilització de dades.

Respecte la directiva de l'any 2003 i, entre d'altres, afegeix definicions de formats llegibles per màquina i formats oberts; estableix que els estats membres han de facilitar els seus documents, sempre que sigui possible i apropiat, en formats oberts i llegibles per màquina i conjuntament amb les seves metadades; especifica que quan s'apliqui una tarifa per a la reutilització dels documents aquesta s'ha de limitar als costos marginals de reproducció, posada a disposició de qualsevol ciutadà i difusió d'aquests documents; i obliga als Estats membres a adoptar i publicar les disposicions legals, reglamentàries i administratives necessàries per donar compliment al que estableix aquesta directiva abans del 18 de juliol de l'any 2015.

2.2.2 Lleis estatals sobre reutilització de la informació del sector públic

A l'estat espanyol la reutilització de la informació en el sector públic es regula per la Llei 37/2007⁴, que té en compte que la informació generada des de les instàncies públiques és de gran interès i pot tenir molta potencialitat per les empreses, que poden contribuir al creixement econòmic i a la creació de llocs de treball, i per als ciutadans com a element de transparència i guia per a la participació democràtica. Aquesta última norma disposa un marc general mínim per a les condicions de reutilització dels documents del sector públic.

Arran de la publicació de la directiva europea de l'any 2013, l'estat espanyol aprova, el 9 de juliol de 2015, la Llei 18/2015⁵ per la que es modifica la Llei 37/2007. Dita llei incorpora a l'ordenament jurídic espanyol els canvis introduïts per la directiva europea sobre reutilització de la informació del sector públic de l'any 2013.

La llei obliga a les administracions i organismes públics a autoritzar la reutilització dels seus documents, amb excepció d'aquells que siguin d'accés restringit. Un aspecte important pels arxius públics és que, com a novetat, la llei amplia el seu àmbit d'aplicació a biblioteques, arxius i museus, ja que es té en compte l'important volum de recursos d'informació de què disposen aquests ens i els projectes de digitalització que estan duent a terme.

A banda, la llei també millora la regulació dels formats a utilitzar per fer disponible la informació del sector públic, promovent l'ús de formats oberts i llegibles per màquina (juntament amb les seves metadades), quedant recollida la seva definició en la mateixa llei.

⁴ [Llei 37/2007](#), de 16 de novembre, sobre reutilització de la informació del sector públic.

⁵ [Llei 18/2015](#), de 9 de juliol, per la que es modifica la Llei 37/2007 sobre reutilització de la informació del sector públic.

La llei també incorpora en el càlcul del règim de tarifes per la reutilització de documents el principi de costos marginals contemplats en la directiva europea, a excepció dels arxius, biblioteques i museus i centres on part del seu pressupost depenguin, en part, de la seva capacitat de generar ingressos, institucions que podran sobrepassar el límit d'aquests costos marginals.

Finalment, la llei incorpora l'obligació prevista en la directiva europea de fomentar l'ús de llicències obertes de manera que la reutilització de la informació del sector públic plantegi les mínimes restriccions possibles.

En definitiva, la legislació europea i espanyola relativa a la reutilització de la informació del sector públic fomenta posar a l'abast dels ciutadans informació de caràcter públic de manera assequible (tant tècnicament com econòmica), permet que aquesta informació es pugui fer servir per finalitats comercials i estableix que la informació s'ofereixi, sempre que sigui possible i apropiat, en formats que en facilitin la seva reutilització (formats oberts).

2.2.3 Altres lleis de l'estat espanyol relacionades amb la RISD

A l'estat espanyol es va aprovar, l'any 2010, una normativa addicional a la Llei 37/2007, amb la voluntat de regular l'Esquema Nacional d'Interoperabilitat en l'àmbit de l'Administració Electrònica⁶. El decret defineix les normes a través de les quals queda garantida la interoperabilitat entre sistemes i estableix els criteris i les recomanacions que permetin i afavoreixin la interoperabilitat a les administracions públiques des d'una perspectiva global i no fragmentària.

Una altra llei estretament vinculada amb la normativa referent a la reutilització de la informació del sector públic és la Llei 19/2013 de Transparència, Accés a la Informació Pública i Bon Govern⁷. A l'article 5.4 d'aquesta llei s'indica que la informació sotmesa a les obligacions de transparència serà publicada a les seues electròniques o pàgines web de les diferents institucions i d'una manera clara, estructurada i entenedora pels interessats i, preferiblement, en formats reutilitzables. D'altra banda, a l'article 11.c s'especifica que la informació publicada als portals de transparència s'haurà d'adequar als principis d'accessibilitat, interoperabilitat i reutilització.

Com es pot apreciar, la llei de transparència estatal demana que la informació es faciliti preferentment en formats reutilitzables (article 5.4). La llei catalana de transparència⁸, però, és molt més taxativa que la norma estatal (MATAS i BALAGUER, J., 2016), establint que tota la informació del sector públic susceptible de ser obtinguda pels ciutadans, tant la que

⁶ [Reial Decret 4/2010](#), de 8 de gener, pel qual es regula l'Esquema Nacional d'Interoperabilitat en l'àmbit de l'Administració Electrònica.

⁷ [Llei 19/2013](#), de 9 de desembre, de Transparència, Accés a la Informació Pública i Bon Govern.

⁸ [Llei 19/2014](#), de 29 de desembre, de Transparència, Accés a la Informació Pública i Bon Govern.

s'ha de difondre per publicitat activa com la que es lliuri a demanda específica d'un ciutadà, s'ha de lliurar en suport o mitjans informàtics que en permetin la reutilització (art. 5, 6.e i 16).

2.3 Plataformes de publicació de dades obertes

Les plataformes de publicació de dades són una peça clau a l'hora de donar el suport tecnològic necessari en una estratègia de dades obertes per part de qualsevol administració o organisme públic. Es pot optar per desenvolupar una plataforma personalitzada ajustada a les necessitats de la institució, o bé optar per productes que ja estan preparats per al seu ús immediat. En qualsevol cas, la plataforma on es publiquen les dades de qualsevol administració hauria d'incloure:

1. El catàleg de dades: element principal de qualsevol plataforma de publicació de dades. El catàleg hauria de comptar amb:
 - Suport per a metadades: són elements que proporcionen informació sobre cadascun dels *datasets* que es publiquen i que permeten entendre la finalitat de les dades i les seves característiques.
 - Compatible amb estàndards: per tal que les dades publicades siguin fàcilment accessibles i possibilitin l'anàlisi i l'intercanvi de dades.
 - Suport avançat per Linked Open Data: que possibiliti que a mitjà i a llarg termini la plataforma pugui evolucionar en l'ús d'aquesta tecnologia i es puguin publicar dades en format *Linked Data*.
2. Serveis de dades: eines que simplifiquin l'accés a les dades i en millorin l'experiència d'ús. Entre aquestes es podrien destacar:
 - Accés directe a les dades: ja sigui a través d'enllaços externs als conjunts de dades allotjats pels propis gestors, mitjançant descàrregues des de la pròpia plataforma, o proporcionant API's (Interfícies de Programació d'Aplicacions) d'accés obert per tal de facilitar el treball dels reutilitzadors que busquen crear nous serveis en base a les dades.
 - Cerca i classificació: presentant resultats rellevants combinant diferents tipus de cerques.
 - Actualització: disposar dels mecanismes necessaris per gestionar els canvis o correccions que es puguin produir sobre les dades. A més, les dades s'han de poder referenciar de forma estable (que no es trenquin les referències i enllaços a les dades) i l'actualització d'aquestes haurien de ser "en viu", és a dir, que si es produeixen canvis en les bases de dades d'origen aquests es vegin reflectits en les dades que hi ha en el portal de dades obertes.

- Visualització: opció recomanable per tal de fer més amigables les dades exposades a la plataforma (geolocalització o comparació de dades existents).
 - Galeria d'aplicacions: facilitar un enllaç senzill i ràpid a les aplicacions que s'han desenvolupat a partir de les dades publicades.
3. Serveis informatius: amb la voluntat d'arribar a un públic objectiu i captar el seu interès pot ser interessant posar informació com:
- Descripció de la iniciativa: introducció al concepte de dades obertes, motivacions i característiques.
 - Notícies, novetats i actualitzacions.
 - Documentació i ajuda per l'usuari: especificar els estàndards utilitzats i documentar-los per tal que qualsevol usuari hi pugui tenir accés.
 - Termes d'ús i llicència: per tal que els usuaris puguin identificar i localitzar fàcilment quines són les condicions d'ús i les llicències aplicades a les dades.
 - Estadístiques d'ús de la plataforma.
4. Serveis per a la comunitat: una de les peces clau en projectes d'obertura de dades serà el de poder comptar amb el suport d'una àmplia comunitat de reutilitzadors i usuaris de les dades. És per això que és important poder comptar amb les eines adequades per mantenir un diàleg permanent amb aquesta comunitat i potenciar-la activament. Amb tot, hi hauria d'haver espai per:
- Comentaris, idees i suggeriments.
 - Peticions de dades: preguntar obertament als ciutadans quins tipus de dades volen.
 - Integració amb les xarxes socials: per tal d'arribar al màxim d'audiència possible.

2.3.1 Principals plataformes disponibles al mercat

No totes les administracions tenen capacitat per desenvolupar les seves pròpies plataformes de dades obertes. Una alternativa molt vàlida és la de recórrer a plataformes *open source* creades per tercers.

La idea de programari *open source* o de codi obert se centra en la premissa que al compartir codi, el programa resultant tendeix a ser d'una qualitat superior a la del software propietari, a banda de ser més ètic, al ser possible la seva compartició. Amb tot, el software de codi obert ofereix accés al codi font (per modificar-lo, corregir-lo o afegir-hi més prestacions), és gratuït (ja que el software es pot obtenir lliurement) i evita els monopolis del software propietari. Aquest moviment ha facilitat enormement l'electrònica de consum, l'accés a la tecnologia en països en vies de desenvolupament i la generació de nous dispositius, negocis i mercats amb actors tant importants com el sistema operatiu GNU/Linux o el seu derivat Android o

programes com LibreOffice o Firefox, que han permès reduir els costos i presentar-se com alternatives de qualitat, i fins i tot i en alguns casos, superiors al software propietari.

Existeix una àmplia comunitat que manté i desenvolupa programari de codi obert, entre la que s'hi troba l'organització sense ànim de lucre [Open Source Initiative \(OSI\)](#), fundada l'any 1998 per Bruce Perens i Eric S. Raymond i dedicada a promocionar aquest tipus de programari.

La opció de recórrer a plataformes de codi obert té l'avantatge de simplificar el procés però l'inconvenient de no poder personalitzar i adaptar del tot la plataforma a les necessitats concretes de cada administració.

Algunes d'aquestes plataformes són les que es descriuen a continuació:

- CKAN (*Comprehensive Knowledge Archive Network*): per les seves prestacions i per les possibilitats que ofereix, és una de les plataformes *open source* més importants per implementar portals de dades obertes. És una solució completa que permet l'accés a les dades obertes i que proporciona eines que agiliten la publicació, compartició, cerca i ús d'aquestes. La plataforma està orientada a editors o publicadors de dades obertes (governos, organitzacions, empreses) i potencia la participació i reutilització de les dades obertes que s'hi publiquen. Entre les característiques més rellevants es podrien mencionar:
 - Publicar i gestionar conjunts de dades: permet importar *datasets* de forma directa o a través d'una interfície web, filtrar per etiquetes, conèixer l'historial de canvis en les dades o veure les versions antigues.
 - Emmagatzemar i gestionar les dades: permet guardar les dades en brut juntament amb les seves metadades, visualitzar-les com dades estructurades amb taules interactives, gràfics o mapes, gestionar les dades per categories i permisos o recol·lectar les dades automàticament d'altres repositoris.
 - Permet federar xarxes amb altres nodes CKAN o construir comunitats amb extensions que permetin als usuaris comentar i seguir els conjunts de dades.
 - Personalitzar: permet una personalització avançada, tant del catàleg com de la seva funcionalitat, a través dels seus mecanismes per al desenvolupament de temes i extensions.

CKAN està disponible de forma gratuïta com a software lliure però els seus creadors ofereixen serveis addicionals de desenvolupament, allotjament, suport i formació.

Moltes institucions utilitzen la plataforma per publicar les seves dades en format obert. Alguns exemples són el portal de dades obertes del [Regne Unit](#), [Itàlia](#), [Noruega](#), [Buenos Aires](#), [Santander](#), [Canàries](#) o [Aragó](#).

- DKAN: és un clon de CKAN però basada en la plataforma [Drupal](#) i que permet la publicació i gestió de dades obertes. Ofereix totes les prestacions que possibiliten la catalogació, publicació i visualització de dades obertes. Entre les seves característiques principals es poden enumerar:
 - Publicació de dades a través de la seva importació directa o a través de l'API.
 - Personalitzar les metadades.
 - Personalitzar el seu aspecte.
 - Emmagatzemar les dades en el propi *site* o externament.
 - Mostrar l'historial de modificacions.
 - Suport per *linked data*.

Hi ha altres alternatives *open source* força interessants que s'enumeren a continuació:

- data.gov.uk. To Go: kit preparat pel govern britànic per posar a disposició del públic en general la seva plataforma de publicació de dades de manera que qui vulgui pugui desplegar una plataforma similar preocupant-se només d'adaptar l'aparença externa final.
- datos.gob.es: és la plataforma utilitzada per la iniciativa de dades obertes de l'estat espanyol i que està disponible en codi obert.

Apart de les plataformes acabades de citar, existeixen al mercat plataformes de software propietari o sota llicència. Algunes de les més destacades són [OGoov](#), [Socrata](#), [Junar](#) o [OpenDataSoft](#).

2.4 Iniciatives governamentals d'obertura de dades

Organitzacions governamentals d'arreu del món han desenvolupat portals a Internet on publiquen dades de caràcter públic amb l'objectiu de fomentar l'ús i la reutilització de la informació procedent de l'administració. En la majoria de casos les dades estan agrupades en catàlegs de dades i estan disponibles en diferents formats.

La Fundació CTIC, organització que lidera la iniciativa mundial de l'*egovernment* del W3C (*World Wide Web Consortium*) i que desenvolupa estratègies per millorar els nivells de transparència del sector públic a través de la publicació de dades en format estàndard, obertes i accessibles, ha realitzat un mapa interactiu ([Open Data World Map](#)) on es mostren les diferents iniciatives d'*open data* per països i el nivell d'obertura de dades segons la classificació feta per Tim Berners-Lee (3 estrelles per dades en formats oberts; 4 estrelles si les dades estan en RDF; i 5 quan són Linked Data).

Així, i a nivell mundial, es podrien destacar els portals de dades obertes del Govern Federal dels Estats Units d'Amèrica (<http://www.data.gov/catalog>), amb 183.444 conjunts de dades;

el del govern australià (<https://www.data.gov.au/>), amb 8.170 *datasets* publicats; o els del Banc Mundial (<http://data.worldbank.org/>) i les Nacions Unides (<http://hdr.undp.org/en/data>).

A nivell europeu destacarien el portal de dades obertes del Regne Unit (<https://data.gov.uk/>), amb 38.933 *datasets*, o el Portal de Dades Obertes d'Europa (<https://data.europa.eu/euodp/es/data>), punt d'accés únic a una gran varietat de dades de les Institucions europees i altres organismes de la UE. En aquest últim portal també s'hi poden trobar relacionats els enllaços que porten a tots els portals de dades obertes dels estats membres de la Unió.

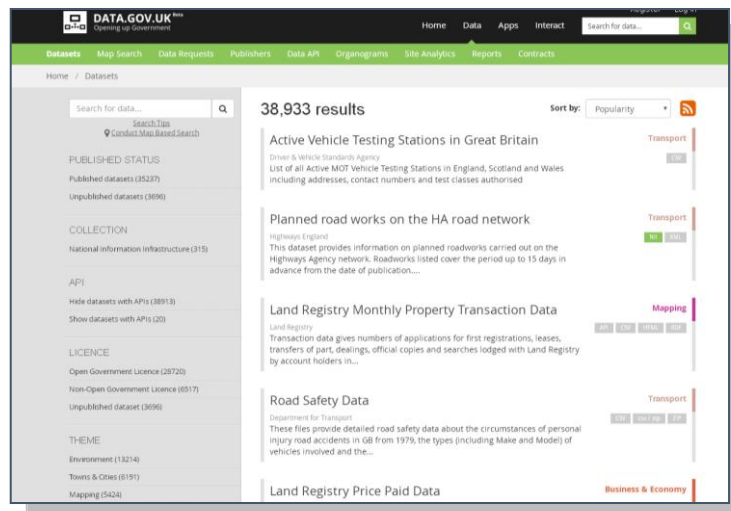


Figura 4. Portal de dades obertes del Regne Unit. Font: <https://data.gov.uk/>



Figura 5. Portal de dades obertes de la Unió Europea. Font: <https://data.europa.eu/euodp/es/data/>

Una eina interessant per veure com està evolucionant el moviment *open data* al món i l'impacte que desperta a la societat i en els negocis, és [l'Open Data Index](http://index.okfn.org/). Elaborat per la *Open Knowledge International* (coneguda fins l'any 2014 com *Open Knowledge Foundation*), l'índex permet mesurar i comparar l'obertura de dades a nivell mundial, conèixer-les i facilitar-ne la seva cerca. Aquest índex va néixer com una necessitat degut al compromís adquirit per molts governs d'obrir les seves dades i per avaluar si aquests últims estaven realment aportant dades als portals creats per aquesta finalitat o només feien declaracions de bones intencions. L'índex permet comparar l'estat de dades obertes entre països i mesurar el progrés en polítiques d'obertura de dades per cada país.

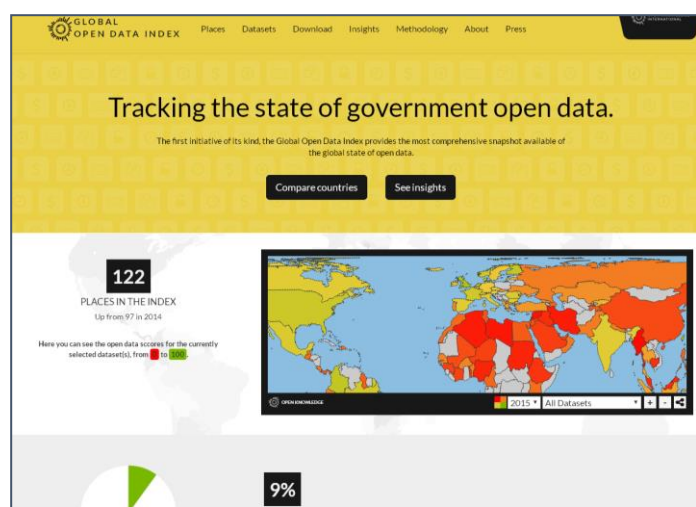


Figura 6. Open Data Index (ODI). Impacte de les dades obertes al món. Font: <http://index.okfn.org/>

2.4.1 Portals de dades obertes estatals

A nivell de l'estat espanyol, el Govern Central de Madrid aprova l'any 2005, i emmarcat dins del *Plan Avanza*, la primera línia d'actuació a nivell estatal dedicada específicament a impulsar l'obertura d'informació del sector públic. Quatre anys més tard, l'any 2009, neix el [Projecte Aporta](#), amb l'objectiu de fomentar una cultura favorable a l'obertura de dades públiques, facilitar que les administracions la puguin dur a terme i impulsar el mercat de la reutilització de la informació pública. L'any 2010, i impulsat pel Ministeri d'Indústria, Energia i Turisme i el Ministeri d'Hisenda i Administracions Públiques, el Projecte Aporta posa en marxa el Catàleg d'Informació Pública, el primer accés únic a nivell estatal a les fonts de dades disponibles del sector públic estatal.

Les seves línies bàsiques d'actuació són:

1. Enriquiment del catàleg d'informació pública existent a datos.gob.es, punt de trobada de la informació pública reutilitzable i publicada per diferents organismes a nivell estatal, regional i local.
2. Establiment del marc legal de reutilització de la informació a nivell estatal: Llei 37/2007, reial decret 1495/2011.
3. Elaboració d'una norma tècnica que estableix les condicions comunes sobre selecció, identificació, descripció, condicions d'ús i posada a disposició de conjunts de dades.
4. Assessorament a les administracions i organismes públics per tal d'assegurar el compliment de la normativa vigent en matèria de reutilització de la informació pública.

A través de la pàgina web del Projecte Aporta es pot accedir a més d'11.000 catàlegs de dades de diferents administracions públiques de l'estat espanyol, es pot visualitzar en un mapa les diferents iniciatives de dades obertes que s'estan desenvolupant a Espanya i es pot accedir a diferents aplicacions desenvolupades a partir de la reutilització de la informació del sector públic.

Tal i com es mostra l'informe elaborat per la fundació Orange l'any 2014 (FUNDACIÓN ORANGE, 2014), i on es recollia el posicionament de les comunitats autònomes i els seus ajuntament més importants en relació a la reutilització de la informació del sector públic, la majoria de comunitats disposen d'un portal de dades obertes i, entre les més avançades, s'hi troben les del País Basc, Catalunya, Aragó i Galícia.

Aquest mateix informe revelava com les principals finalitats que busquen les comunitats autònomes i ajuntaments en el moment d'obrir les seves dades a la ciutadania són augmentar la percepció de transparència que tenen els ciutadans respecte l'administració, fomentar la innovació i la competitivitat, prestar nous serveis amb valor afegit i millorar la qualitat dels serveis que s'ofereixen als ciutadans.

2.4.2 El portal de dades obertes de la Generalitat de Catalunya

Tal i com mostrava l'informe de la Fundació Orange acabat de citar, la Generalitat de Catalunya era, l'any 2014, una de les institucions de l'estat espanyol que publicava un major nombre de conjunts de dades en el seu portal de dades obertes (1.427 *datasets*).

Actualment el [portal de dades obertes de la Generalitat de Catalunya](http://portal.de.dades.obertes.de.la.Generalitat.de.Catalunya) té un catàleg amb 1.490 conjunts de dades. El portal, gestionat des de la Secretaria de Transparència i Govern Obert del Departament d'Afers i Relacions Institucionals i Exteriors i Transparència, publica dades procedents de diferents organismes públics de la Generalitat i té la voluntat d'agrupar totes les iniciatives de dades obertes de la Generalitat en un únic catàleg.

A banda de la publicació d'aquestes dades, el portal ofereix informació sobre el concepte de dades obertes, els seus formats més habituals, facilita enllaços als diferents portals de dades obertes que hi ha arreu del territori espanyol i a nivell europeu i mundial, mostra aplicacions que s'han desenvolupat amb les dades que s'han obert i facilitat a la ciutadania i inclou un apartat de col·laboració en el projecte *open data* de la Generalitat de Catalunya a partir del qual qualsevol ciutadà pot exposar les dades que creu que haurien d'obrir-se i aportar idees per tal que es desenvolupin noves aplicacions amb les dades que es faciliten al portal.

2.5 Experiències d'institucions arxivístiques en obertura de dades

Tot i que el moviment *open data* en el món dels arxius és encara incipient, ja hi ha algunes institucions arxivístiques que han donat les primeres passes en l'alliberament de les seves dades i estan començant a oferir els seus inventaris i catàlegs en formats oberts per tal que la informació continguda en aquests pugui ser reutilitzada per qualsevol persona o institució que hi estigui interessada.

Tot seguit es passen a enumerar algunes de les iniciatives d'aquest tipus que estan duent a terme centres arxivístics d'arreu del món.

2.5.1 Iniciatives internacionals

A nivell mundial, algunes de les iniciatives d'obertura de dades en arxius són les que s'enumeren a continuació.

2.5.1.1 NARA (National Archives and Records Administration dels Estats Units d'Amèrica)

El NARA és una agència independent adscrita al Govern Federal dels Estats Units d'Amèrica, que protegeix i documenta els registres governamentals i històrics. La majoria dels documents que estan sota la seva custòdia són de domini públic, ja que els treballs del Govern Federal estan exclosos de la protecció dels drets d'autor. La seva missió és la de proporcionar accés públic als documents del Govern Nacional dels Estats Units d'Amèrica.

En un dels apartats de la seva pàgina web es descriuen les seves polítiques d'*open data*, que estan incloses en les estratègies de govern digital i de l'*open government* dins dels mateixos arxius nacionals (<http://www.archives.gov/data/>).

Des de l'any 2013 els Arxius Nacionals nord-americans estan treballant per implementar polítiques de dades obertes tal i com estableix l'Ordre Executiva EO13642⁹ i posen a

⁹ [Ordre Executiva EO 13642](#) signada pel president dels Estats Units d'Amèrica, Barack Obama, el 9 de maig de 2013, que ordena que totes les agències nord-americanes hauran d'implementar els requeriments de l'*open data* i la informació que generin haurà de ser llegible per ordinador i en formats oberts per defecte (a excepció d'aquella que no pugui ser consultable per raons de

disposició dels ciutadans que ho desitgin conjunts de *datasets* en diversos formats (xml, csv, pdf i html). Es pot accedir a tot l'inventari de dades obertes dels NARA en format json o accedir de manera individual a cadascun dels conjunts de dades obertes publicats. Aquests conjunts de dades obertes estan disponibles al portal de dades obertes dels Estats Units d'Amèrica (<https://www.data.gov/>).

Entrant al portal de dades obertes nord-americà i cercant l'organització NARA es pot comprovar com hi ha disponibles un total de 49 conjunts de dades. El portal permet ordenar els resultats per rellevància, nom ascendent, nom descendent, data de la última modificació, popularitat i data d'agregació al portal. També es permet la tria per categories o formats dels arxius. Val a dir que hi ha publicats conjunts de dades que són accessibles per tothom com aquells que no ho són. En aquest últim cas, al no ser accessibles, no es poden descarregar els conjunts de dades en cap tipus de fitxer i només s'ofereix informació bàsica del recurs (a través de les metadades que permet complimentar CKAN – plataforma utilitzada en aquest portal). En tots els recursos es poden descarregar les metadades que descriuen el conjunt en format json. Aquestes metadades donen informació sobre el tipus d'accés i llicència d'ús de les dades, les dates de creació del recurs i d'actualització, els *tags* assignats al recurs i altra informació addicional com ara el tipus de recurs, l'entitat que el publica, la persona que el manté i la llengua del recurs, entre d'altres.

Entre els *datasets*, n'hi ha que estan en format html i que l'únic que fan és donar informació bàsica sobre alguna de les funcions que desenvolupen els arxius nacionals. En aquests casos només es publica un enllaç que redirecciona l'usuari a l'apartat específic de la pàgina web de la institució que descriu el conjunt. Un exemple és el conjunt de dades que porta per nom [Archives.gov](#), on es dona informació del portal web dels NARA.

D'altra banda, hi ha conjunts de dades que es poden descarregar en formats oberts (com csv). Un exemple és el [Amending America: Proposed Amendments to the United States Constitution, 1787 to 2014](#). El conjunt de dades proporciona informació sobre més d'11.000 propostes d'esmenes constitucionals introduïdes al Congrés dels Estats Units des de 1787 fins al 2014. Es pot descarregar el conjunt del fitxer en format csv o previsualitzar-lo.

A banda de l'aportació d'aquests 49 conjunts de dades al portal de dades obertes del govern nord-americà, un aspecte molt interessant que ofereix el [catàleg en línia dels NARA](#) és la possibilitat d'exportar els resultats que s'obtenen a partir de qualsevol cerca que es faci en diferents formats, molts dels quals oberts i aptes per la seva reutilització (csv, json o xml). Es poden exportar tots els resultats obtinguts d'una cerca o seleccionant-los un a un. A la

vegada, es pot optar per descarregar la descripció completa de qualsevol unitat de descripció o un resum.

2.5.1.2 LAC (*Library and Archives Canada*)

La *Library and Archives Canada* és la institució federal canadenca encarregada d'ingressar, preservar i fer accessible el patrimoni documental del Canadà.

L'any 2011 aquesta institució es va unir a la iniciativa de govern obert del Govern del Canadà, publicant al portal de dades obertes del país dos conjunts de dades: [la relació](#) de les més de 600.000 persones que es van allistar a les Forces Expedicionàries Canadenques durant la Primera Guerra Mundial com a soldats, infermeres o capellans i [la descripció](#) de més de 40.000 mapes, plànols i gràfics que formen part del patrimoni cartogràfic canadenc.

Des del 2011 la LAC ha anat alliberant els conjunts de dades que són més consultats pels seus usuaris, fins arribar al total de [84 datasets](#) que hi ha publicats actualment al portal de dades obertes del Canadà. A la pàgina web de la LAC s'hi pot trobar un accés directe que enllaça als conjunts de dades que la institució té publicades al portal.

Com en la majoria de portals de dades obertes, en el del Canadà es poden cercar conjunts de dades segons l'organisme que els ha publicat, el tema del *dataset* o el format del fitxer. Si s'accedeix a cada recurs, es pot obtenir (mitjançant les metadades) informació relativa a la data de publicació del conjunt, la data de publicació, l'enllaç al recurs de la institució que el publica i les seves dates extremes.

Com s'acaba de comentar, la Biblioteca i Arxius del Canadà disposa de 84 conjunts de dades, entre els que hi destaquen els dos conjunts citats anteriorment i un [conjunt](#) de 1.000 documents relacionats amb personal de les Forces Expedicionàries Canadenques durant la Primera Guerra Mundial publicats com a Linked Open Data (LOD).

2.5.1.3 TNA (*The National Archives del Regne Unit*)

Els Arxius Nacionals del Regne Unit és un organisme governamental britànic constituït l'any 2003 per ser un arxiu nacional dels països d'Anglaterra, Gal·les i Irlanda del Nord i on s'hi custodien documents amb més de 1.000 anys d'història.

Entre els anys 2003 i 2006 quatre òrgans del govern britànic (cadascun d'ells especialitzats en aspectes particulars de la gestió de la informació), es van unir per formar una única organització: els Arxius Nacionals del Regne Unit. Aquests quatre òrgans eren:

1. La *Public Record Office*: dedicada a la preservació dels registres públics clau i fer-los accessibles als investigadors.
2. La *Royal Commission on Historical Manuscripts*: amb funcions relacionades amb la preservació de documents històrics de caràcter confidencial.

3. *Her Majesty's Stationery Office*: titular dels drets d'autor de la Corona Britànica i impressora oficial de totes les lleis del parlament britànic des de l'any 1889.
4. La *Office of Public Sector Information*: creada l'any 2005 per promoure la reutilització de la informació produïda i recollida per les organitzacions del sector públic.

Els Arxius Nacionals britànics són els encarregats de fixar les normes relacionades amb la gestió de la informació de tot el Regne Unit i treballen amb la voluntat de proporcionar el marc pràctic adequat per fomentar la reutilització de la informació del sector públic britànic.

Les descripcions dels documents custodiats pels Arxius Nacionals del Regne Unit es poden consultar en línia a través de la seva pàgina web. Són un total de 32 milions de descripcions de documents custodiats en més de 2.500 centres arxivístics d'arreu del territori britànic. Tot el contingut que es publica en aquesta pàgina web està disponible sota la *Open Government Licence v3.0*¹⁰, excepte quan s'indica el contrari. Això fa que la majoria dels resultats obtinguts mitjançant qualsevol cerca que es realitzi es puguin exportar en format oberts (csv, html o xml).

Així per exemple, si es fa una cerca avançada en el [catàleg en línia dels Arxius Nacionals del Regne Unit](#) i es busquen els documents relacionats amb Barcelona entre els anys 1925 i 1949, s'obtenen 252 descripcions de documents. Els resultats obtinguts es poden filtrar per data del document, fons documental o col·lecció a què pertany, subjectes, nivell de descripció, règim d'accés al document (obert, restringit o retingut) o data en què la descripció es va pujar a la web. Tots els resultats obtinguts es poden exportar i descarregar en csv, html o xml. Aquest procés d'exportació té una limitació: només es poden descarregar en formats oberts les 1.000 primeres descripcions dels resultats obtinguts. Val a dir que el cercador que ofereixen els TNA als seus usuaris és realment espectacular.

A més de permetre la descàrrega de la majoria de les descripcions dels documents en format oberts, els TNA ofereixen un recurs molt interessant: és el [The National Archives Labs](#). És un portal web pensat per involucrar l'usuari dels Arxius Nacionals del Regne Unit en el desenvolupament de noves idees i maneres de presentar la informació que contenen els instruments de descripció que la institució posa a disposició de l'usuari a través de la seva pàgina web principal. És una mena de banc de proves on enginyers, desenvolupadors i un equip d'investigació dels Arxius Nacionals mostren què es pot fer amb les dades de què disposa l'arxiu i en demana l'opinió als usuaris. També és un lloc que mira de buscar la

¹⁰ La [Open Government Licence for Public Sector Information](#) permet a qualsevol persona o organització copiar, publicar, distribuir, transmetre, adaptar i explotar la informació publicada a la pàgina web dels Arxius Nacionals britànics sense cap tipus de restricció. En resta exclosa la informació que contingui dades personals o estigui protegida per drets de propietat intel·lectual. La llicència està pensada i desenvolupada com una eina per tal que els proveïdors d'informació del sector públic posin a disposició de la ciutadania la informació pública de la que disposen sota una llicència oberta per tal que se'n permeti la seva reutilització.

implicació de nous agents per tal que desenvolupin les seves pròpies aplicacions amb les dades de la institució. Els projectes que tenen èxit en aquest banc de proves són pujats a la pàgina web principal dels Arxius Nacionals. El que es pot trobar, bàsicament, en aquesta pàgina web és:

1. [Datasets](#) en formats oberts descarregats de la pàgina web principal dels TNA i que l'usuari es pot carregar a seu ordinador. Entre aquests s'hi poden trobar:
 - Relació de la documentació rebuda pels TNA durant els últims deu anys, incloent informació sobre el departament d'origen, l'any d'ingrés, la referència del catàleg i l'espai que ocupen a les prestatgeries de l'arxiu.
 - Correspondència antiga (1175-1538): catàleg d'una selecció de cartes dels segles XII al XVI que proporciona informació sobre els escriptors, els receptors i el tema de la correspondència.
2. Visualització de proves fetes amb dades procedents dels instruments de descripció disponibles a la pàgina web dels TNA. Alguns exemples són:
 - [Visualització](#) de l'evolució dels diferents departaments encarregats dels afers exteriors britànics des de l'any 1945 fins a l'actualitat. Els noms dels diferents departaments que consten al *timeline* permeten enllaçar amb la descripció del fons a la web dels TNA.
 - [Collections on a map](#): permet veure sobre un mapa diferents unitats de descripció de diferents fons dels TNA i accedir a la descripció de cada document.
 - [Cabinet papers keywords](#): aplicació que mostra els noms dels membres dels Consells de Ministres britànics entre els anys 1917 i 1980 i les 25 paraules clau de totes les reunions pel període indicat. Cada entrada està vinculada a les actes de les reunions (que es poden descarregar gratuïtament). Permet veure les paraules més utilitzades per un rang màxim de 10 anys.

A banda de les possibilitats que ofereix el catàleg en línia dels TNA en relació a la reutilització de la informació, la mateixa pàgina web té un apartat destinat a la [transparència](#). La informació que consta en aquest apartat és la mateixa (o similar) a la que els Arxius Nacionals Britànics aporten al portat de dades obertes del govern britànic (<https://data.gov.uk/>), on hi ha un total de [39 datasets publicats per la institució](#). Aquí els TNA ofereixen, entre d'altra, informació relativa a:

1. Energia, medi ambient i sostenibilitat: dades sobre l'ús de l'electricitat a l'edifici de l'arxiu nacional.
2. Despeses: en què gasten els diners els Arxius Nacionals: informació relativa a les fonts de finançament, ingressos que es generen, informació sobre els pagaments als

proveïdors que superen les 10.000 lliures, salaris, drets de pensió i altres beneficis fiscals dels funcionaris d'alt nivell que treballen als Arxius Nacionals o les despeses dels directors dels diferents centres d'arxius britànics.

3. Quantitat de documents servits als usuaris (en línia o presencialment) i nivells de satisfacció d'aquests en relació al servei ofert per la institució.
4. Quantitat de documents que es digitalitzen i quants d'aquests es poden a disposició dels usuaris en línia.

2.5.1.4 SAN (*Sistema Archivistico Nazionale*)

Inaugurat el 17 de desembre de 2011, el [Portal SAN](#) és l'agregador nacional italià dels recursos arxivístics del país transalpí i constitueix un punt d'accés integrat al patrimoni arxivístic italià. Ofereix descripcions i accés a recursos arxivístics de naturalesa heterogènia, estatals i no estatals, públics o privats. La responsabilitat del manteniment del portal recau sobre l'*Istituto Centrale per gli Archivi* (ICAR), organisme d'estudi i recerca aplicada depenent de la *Direzione generale per Gli Archivi* (Direcció General dels Arxius Italians). Actualment formen part del SAN un total de 21 sistemes d'arxius amb un total de 553.214 recursos arxivístics i digitals (CHIARETTI, 2014). És, en definitiva, un repositori de dades arxivístiques on hi ha informació de conjunts arxivístics, productors, instruments de descripció i conservadors. Els sistemes adherits al SAN envien les dades en base a un format d'intercanvi (*xml schema*), definit a nivell central l'any 2009.

A principis de l'any 2013 s'inicià un projecte per desenvolupar i publicar les dades del SAN en format de dades obertes enllaçades (LOD). Actualment es troba publicada la versió 1.1 de l'ontologia, en format OWL (*Ontology Web Language*), capaç de representar l'esquema conceptual expressat per les dades. Juntament amb l'ontologia hi ha disponibles totes les dades (5.312.474 tripletes i 128 *datasets*) per ser descarregades i reutilitzades. A més, amb la voluntat d'estendre el model conceptual del SAN i la seva estructura d'informació, s'han formalitzat, a través del llenguatge SKOS¹¹, vocabularis controlats que permeten introduir noves entitats i noves possibilitats d'accés.

Donar accés al patrimoni arxivístic del SAN com a *Linked Open Data* pretén aconseguir la interoperabilitat tecnològica i semàntica entre els diferents sistemes arxivístics italians i permet oferir serveis d'ús avançat als usuaris. També es busca augmentar la visibilitat de les dades arxivístiques i permetre noves connexions entre subjectes, persones, organitzacions i

¹¹ SKOS (*Simple Knowledge Organization System*) és una iniciativa del W3C en forma d'aplicació RDF que proporciona un model per representar l'estructura bàsica i el contingut d'esquemes conceptuals com ara llistes d'encapçalaments de matèria, taxonomies, esquemes de classificació, tesaurus i qualsevol altre tipus de vocabulari controlat.

llocs amb la finalitat de promoure la investigació interdisciplinària i enriquir el coneixement històric.

2.5.2 Iniciatives arreu de l'estat espanyol

A nivell estatal espanyol ja hi ha algunes entitats arxivístiques que estan començant a treballar i a oferir els seus instruments de descripció en formats que puguin ser reutilitzables. Entre les més destacades s'hi poden trobar les que s'expliciten en els següents apartats.

2.5.2.1 L'Arxiu d'Arganda del Rey

Una de les experiències més emblemàtiques d'obertura de dades a l'estat espanyol és el cas de l'Arxiu d'Arganda del Rey. L'Arxiu conserva els documents produïts i rebuts per l'administració local des de fa més de 800 anys, un notable fons documental en diferents suports.

L'any 1999 va ser el primer arxiu de l'estat espanyol en crear un catàleg obert que donés accés a tothom qui volgués consultar els seus registres descriptius i a les seves imatges associades. Els anys 2006 i 2012 va renovar la seva pàgina web apostant, en aquest últim cas, per la filosofia *open* i les utilitats de la web 3.0.

El seu actual sistema de gestió està basat en el concepte d'arxius oberts utilitzant el protocol OAI-PMH (*Open Archives Initiative – Protocol for Metadata Harvesting*) que permet la lliure descàrrega, distribució i enllaçat de dades normalitzades i estructurades. El seu major potencial és que permet la recol·lecció automatitzada de fitxers en format xml amb esquemes normalitzats, facilitant la seva posterior reutilització, intercanvi i explotació sense cap mena de barrera financera, tècnica o semàntica.

El sistema es basa en una solució *low cost* que garanteix la interoperabilitat de les dades, seguint la filosofia *open/linked data*, usant *datasets* gestionats en un repositori OAI-PMH desenvolupat íntegrament amb recursos propis i amb software *open source*, amb la qual cosa possibilita la seva transferència tecnològica.

L'any 2015 l'Arxiu d'Arganda del Rey va rebre, per aquest projecte, el premi CNIS (Congrés Nacional d'Innovació i Servei Públic) al millor projecte en transparència i *open data*.

Tota la tecnologia que usa és aplicable a qualsevol dada de caràcter públic i, com s'acaba de comentar, l'arquitectura del sistema gira al voltant del protocol d'arxiu oberts per a la recol·lecció de metadades (OAI-PMH), un repositori OAI-PMH que incrementa exponencialment la disponibilitat, accessibilitat i reutilització dels diferents catàlegs de dades. És un complement a la cerca en el portal web convencional, que segueix existint, i que es diferencia de la opció de descàrrega habitual de fitxers en el fet que la infraestructura tècnica de la OAI obliga a complir uns requisits de normalització per tal de poder accedir als

datasets del seu repositori, un format comú pels fitxers xml (XML-DC¹², XML-EAD¹³ i XML-ESE¹⁴), la codificació i identificació de les seves metadades i, per tant, informació normalitzada que fa que la institució es converteixi en proveïdora de continguts de qualitat que poden ser recol·lectats automàticament per altres portals i agregadors com Hispana¹⁵ o Europeaana¹⁶.



Figura 7. Arquitectura del sistema d'arxiu de l'Arxiu Municipal d'Arganda. Font: [Arxiu Municipal Arganda del Rey](#)

A banda, l'Arxiu d'Arganda del Rey ha vinculat els descriptors de les seves bases de dades a vocabularis controlats com *Geonames*¹⁷ (descriptors geogràfics), *Dbpedia*¹⁸ (matèries) i

¹² Representació del model de metadades *Dublin Core* en llenguatge xml. L'esquema de dades *Dublin Core* és un conjunt d'elements per descriure recursos i que comprèn quinze elements bàsics. Els sistemes basats en *Dublin Core* utilitzen normalment xml i es basen en RDF.

¹³ Representació del codi de descripció arxivística *Encoded Archival Description (EAD)* en llenguatge xml. L'*Encoded Archival Description* és un model de codificació dels inventaris d'arxius basat en el llenguatge xml i que s'ha generalitzat en molts països. El seu desenvolupament va començar a la Universitat de Califòrnia a Berkeley i el seu objectiu era el de crear un estàndard de codificació de la descripció de documents d'arxius que pogués ser interpretada per un ordinador amb independència del programari i maquinari utilitzats, facilitant-ne així les cerques. És un format que contribueix a la reutilització de la informació ja que permet importar i exportar la informació entre diversos sistemes.

¹⁴ Representació del model de metadades *Europeana Semantic Elements* en llenguatge xml. *Europeana Semantic Elements* és el model de metadades adaptat per *Europeana* fins l'any 2011. Establia un conjunt de camps *Dublin Core* juntament amb altres dotze elements específics d'*Europeana*.

¹⁵ Hispana és el porta d'accés a la cultura digital i l'agregador nacional de continguts a Europeana. Reuneix les col·leccions digitals d'arxius, biblioteques i museus de l'estat espanyol. Aquests poden participar a Hispana de tres maneres: donant-se d'alta al directori de projectes de digitalització; aportant els seus registres a Hispana (el requisit és disposar d'un repositori OAI); o aportant les seves col·leccions a Europeana (el repositori OAI ha d'aportar els registres en format ESE o EDM).

¹⁶ El porta Europeana actua com a punt d'accés únic a milions de recursos digitals d'arxius, museus, biblioteques i col·leccions audiovisuals europees.

¹⁷ *Geonames* és una base de dades geogràfica, gratuïta i accessible a través d'Internet sota una llicència *Creative Commons* Reconeixement 3.0. Conté més de 8 milions de noms geogràfics que corresponen a més de 6,5 milions de llocs existents.

VIAF¹⁹ (persones), enriquint aquestes dades, millorant la navegació per la seva pàgina web i fent possible les cerques multilingües. Així, el projecte es pot relacionar amb la web semàntica.

Per la creació del repositori OAI-PMH l'Arxiu d'Arganda del Rey ha utilitzat un software *open source*, Repox²⁰, al que s'hi ha agregat noves funcionalitats que l'arxiu ha considerat necessàries.

Al mateix temps l'Arxiu també ha millorat el sistema de recuperació de la informació del seu catàleg en línia. Així, entre d'altres, es poden fer filtres i cerques sobre els resultats obtinguts en les cerques, geolocalitzar la informació sobre mapes i interaccionar amb els usuaris permetent-los que puguin fer les seves aportacions a través d'un formulari que hi ha visible a cada registre.

2.5.2.2 DARA (Documentos y Archivos de Aragón)

El [portal DARA](#) és un projecte cooperatiu obert a tots els arxius aragonesos per donar accés a través d'internet als arxius i documents de la Comunitat Autònoma d'Aragó. Forma part del SIPCA (Sistema d'Informació del Patrimoni Cultural Aragonès) i està integrat per més de 60 arxius públics i privats que depenen de diferents administracions (Govern d'Aragó, diputacions, comarques i ajuntaments), associacions, fundacions privades i altres institucions).

Amb data 20 de març de 2016, disposava de més de mig milió de referències documentals que descriuen un total de 500 fons documentals i entre les que s'inclouen unes 150.000 fotografies i més de 15.000 volums de protocols notariais.

Per tal de poder ser punt de reunió dels arxius aragonesos, aquests últims treballen amb un format comú, anomenat EDARA (Estructura de Dades dels Arxius Aragonesos), que permet l'intercanvi de registres. L'EDARA està realitzada a partir de l'estructura de la ISAD (G) però amb la vista posada als formats xml, més en concret en l'EAD (Encoded Archival Description).

¹⁸ *Dbpedia* és un projecte, iniciat per investigadors de la Universitat Lliure de Berlín i de la Universitat de Leipzig, l'objectiu del qual és l'extracció automàtica de dades i contingut a partir de la informació creada des del projecte Viquipèdia. Aquesta informació estructurada s'entrellaça amb altres bases de coneixement i es fa disponible a la *Word Wide Web*.

¹⁹ El *Virtual International Authority Field* o *VIAF* és un registre d'autoritats d'abast internacional, dissenyat per poder accedir als principals fitxers de noms i autors en l'alfabet i la llengua preferits per cada usuari. Disposa de més de 20 milions de fitxes i la seva finalitat és la d'enllaçar tots els fitxers nacionals d'autoritats en un únic fitxer d'autoritats en línia.

²⁰ *Repos* és un *framework open source* que permet gestionar espais per a metadades. El seu objectiu és el de proporcionar als socis d'Europeana una solució simple per importar, convertir i exposar les seves dades a través d'OAI-PMH.

Els arxius aragonesos, a través del DARA, posen a disposició de tots els ciutadans les descripcions disponibles al portal en formats de dades obertes (csv, json i xml). I ho fan publicant al portal de dades obertes del Govern d'Aragó, l'[Aragón Open Data](#), el [conjunt de dades corresponents a totes les descripcions que els DARA tenen](#).

Els DARA també tenen grups de treball constituïts per tirar endavant determinats projectes. Entre aquests grups n'hi ha dos que són d'especial interès per la matèria tractada en aquest treball:

1. DARA Open Data: l'objectiu del grup de treball és el de buscar mecanismes d'integració de la informació que hi ha als DARA amb l'Aragón Open Data. Els objectius específics del grup són els de difondre tota la informació dels DARA que estigui disponible en dades obertes i preparar i integrar nous fitxers dels DARA en dades obertes i actualitzar aquesta informació (arxius i fons disponibles, estadístiques d'ús i d'usuaris).
2. OpenDARA – Context i Web Semàntica: grup encarregat d'estudiar com els arxius aragonesos poden utilitzar els seus documents, descripcions, dades i informació en el marc de la web semàntica per oferir els seus recursos, enriquir-los i explotar-los sota aquests estàndards. Analitzen estàndards, aplicacions i projectes per conèixer el panorama actual de la web semàntica en el context dels arxius i de les institucions relacionades (LODLAM²¹).

2.5.2.3 PARES (*Portal de Archivos Españoles*)

PARES és el portal d'arxius de l'estat espanyol que facilita als investigadors i als ciutadans en general l'accés integrat a la xarxa de les bases de dades arxivístiques de vuit grans arxius de titularitat estatal distribuïts per la geografia espanyola (SÁNCHEZ MAIRENA, 2014). Actualment és un repositori per la creació, gestió i difusió dels arxius històrics.

La interoperabilitat amb els grans projectes de difusió europea va incidir en la necessitat de fer evolucionar PARES cap a un sistema d'arxius oberts, iniciant la transformació del portal, l'any 2010, segons la filosofia *Open Archives Initiative – Protocol for Metadata Harvesting* (i que s'explicarà en el punt 2.5.3 del present treball).

Actualment PARES disposa de 8,6 milions registres arxivístics i unes 34 milions d'imatges. El 10 de maig de l'any 2016 [es va presentar la nova versió de PARES](#), la PARES 2.0, que, segons els seus desenvolupadors, avança de manera significativa en la interoperabilitat amb altres sistemes d'informació cultural nacionals i internacionals (com el Portal d'Arxius

²¹ [LODLAM](#) (*Linked Open Data in Libraries, Archives and Museums*): és una xarxa informal i sense fronteres d'aficionats, tècnics, professionals i persones interessades a treballar amb projectes de dades enllaçades en arxius, biblioteques i museus. La seva pàgina web està dissenyada com a lloc central d'intercanvi de recursos i de col·laboració amb d'altres persones interessades.

Europeus²² o Europea) i posa a disposició de les persones que hi estiguin interessades la informació arxivística mitjançant dades obertes enllaçades, possibilitant interrogar al sistema en castellà, anglès, francès, català, eusquera i gallec i amb la possibilitat de navegar entre les unitats descriptives, les autoritats arxivístiques i les seves relacions, els punts d'accés, els objectes digitals i els enllaços a recursos externs.

Val a dir, però, que sembla que la nova versió del portal no està encara operativa per ser consultada ja que al portal que s'accedeix des d'internet no ha variat respecte (el que hauria de ser) el seu predecessor.

2.5.2.4 Arxiu Municipal de Girona

L'Arxiu Municipal de Girona és el servei especialitzat de l'Ajuntament de la ciutat en la gestió, el tractament i la preservació de la documentació, amb independència de la seva data, tipologia i suport, generada pel mateix ajuntament al llarg de la seva existència. Per aquesta finalitat s'estructura en tres unitats d'informació:

1. L'Arxiu Administratiu: acull i gestiona els documents que provenen de les oficines municipals i que són d'utilitat directa tant per a l'administració municipal com per garantir els drets de la ciutadania.
2. L'Arxiu Històric: acull i gestiona els documents que han estat conservats per tal de deixar constància i testimoni de l'actuació municipal al llarg de la història.
3. El Centre de Recerca i Difusió de la Imatge: gestiona i preserva específicament els documents sonors i en imatge produïts per l'Ajuntament de Girona i de fons i col·leccions procedents de particulars, empreses i entitats del municipi.

Coincidint amb l'Open Data Day²³ de l'any 2016, l'Arxiu Municipal de Girona s'hi adherí per primera vegada, amb l'obertura de les dades disponibles sobre el patrimoni documental que s'hi preserva. Sobre el total del fons s'ofereixen en obert:

- Informació bàsica de 125 fons i col·leccions
- 31 inventaris
- 75 catàlegs

Així, s'hi s'accedeix a la pàgina web de l'arxiu, entre els apartats destinats a la consulta de documents (cercador) se n'hi pot trobar un que du per nom "[Guia de fons: dades obertes](#)".

²² [El Portal d'Arxius Europeus](#) proporciona accés a la informació dels materials d'arxiu de diferents països europeus i dona informació d'institucions arxivístiques d'arreu d'Europa. Actualment disposa de prop més de 246 milions de descripcions arxivístiques de més de 6.700 arxius.

²³ L'[Open Data Day](#) és una iniciativa, celebrada cada 5 de març i promoguda per la [Open Knowledge](#), que pretén recopilar accions ciutadanes d'arreu del món orientades a desenvolupar aplicacions, alliberar dades, crear visualitzacions i publicar anàlisis utilitzant dades públiques obertes. L'objectiu és encoratjar els governs locals, regionals i nacionals de tot el món en l'adopció de polítiques d'obertura de les dades que generen en el seu dia a dia.

És aquí on qualsevol persona es pot descarregar en format csv la informació bàsica de tots els fons (guia), els inventaris o els catàlegs. Hi ha la possibilitat de navegar entre els diferents fons i descarregar-ne tots els inventaris i catàlegs per tipus o bé descarregar el que més interressi a l'usuari.

Pel procés d'obertura de dades s'han exportat les dades disponibles a les guies, inventaris i catàlegs en format csv (grau 3 estrelles sobre les 5 en l'escala bàsica de l'*Open Data*). L'estructura de les dades es correspon a la descripció arxivística normalitzada i, consegüentment, la informació dels fons, inventaris i catàlegs està alineada. A la mateixa pàgina web de l'arxiu [s'explicita aquesta estructura](#) i es faciliten instruccions als usuaris de com descarregar les dades i poder treballar amb elles.

Tal i com s'apunta des del mateix arxiu, entre les seves línies de futur, tant de l'Arxiu Municipal de Girona com del Centre de Recerca i Difusió de les Imatges, es preveu la integració de l'obertura de dades en qualsevol futur tractament d'organització i descripció de la documentació, i s'està treballant per oferir tant aviat com sigui possible les dades del 100% del fons.

2.5.3 El protocol OAI-PMH i l'*Open Archive Initiative*

En alguns dels casos acabats d'exposar s'ha citat el protocol OAI-PMH. En aquest apartat s'exposa què és i quines oportunitats ofereix pels arxius.

L'*Open Archive Initiative* és una iniciativa nascuda a finals dels anys 90 que proporciona l'arquitectura i les especificacions tècniques necessàries per tal que productors i distribuïdors de continguts oberts posin a disposició d'agregadors de continguts, conjunts de metadades sobre els documents que emmagatzemen amb l'objectiu de fer-los visibles i accessibles als investigadors. Vindria a ser una proposta d'integració i interoperabilitat entre repositoris d'informació que permet la creació de catàlegs col·lectius a través de l'ús de normes i estàndards internacionals, amb una metodologia senzilla i de fàcil implementació (GÓMEZ DUEÑAS, 2005).

L'arquitectura del sistema s'articula en torn del protocol d'arxius oberts per a la recol·lecció de metadades (OAI-PMH), el qual permet que cada comunitat pugui mantenir esquemes de metadades propis malgrat hagin d'incorporar estàndards com *Dublin Core*, *Europeana Semantic Elements* o *Europeana Data Model*²⁴.

El protocol té tres característiques bàsiques:

²⁴ *Europeana Data Model* és el nou model de metadades d'*Europeana* que està substituint gradualment l'ESE, oferint més possibilitats de cerca i contextualització i informació per l'usuari, i d'interoperabilitat en projectes com l'*Open Linked Data*.

1. Simplicitat: els seus creadors van buscar una fórmula simple que estigués a l'abast de qualsevol potencial implementador.
2. Normalització: basat en estàndards àmpliament utilitzats a internet com són el protocol http (*Hipertext Transfer Protocol*) per a la transmissió de dades i ordres, i xml (*Extended Markup Language*) per a la codificació de les metadades.
3. Recol·lecció: existeixen agregadors de continguts que recullen periòdicament totes o part de les metadades exposades en els diferents OAI per poder desenvolupar serveis de valor afegit.

Hi ha dos classes de participants en el marc OAI-PMH:

1. Proveïdors de dades (*data provider*): són entitats que proporcionen les metadades dels documents que gestionen. Administren sistemes que suporten el protocol OAI-PMH com una forma d'exposar les metadades.
2. Proveïdors de serveis (*service provider*): són recol·lectors o agregadors de continguts i que agafen les dades amb l'objectiu d'incorporar-los valor afegit i oferir-los als usuaris finals. Aquests últims poden ser empreses o sistemes públics de difusió com Hispana o Europea.

La comunicació entre el proveïdor de dades i el proveïdor de serveis es fa utilitzant transaccions http per emetre preguntes i obtenir respostes com ara:

- *Identify*: per recuperar informació sobre el repositori (proveïdor de dades)
- *Listmetadataformats*: per recuperar els formats de metadades disponibles des d'un repositori (normalment en *Dublin Core*).
- *Listidentifiers*: recupera els identificadors o codis de referència dels registres. Té un argument obligatori (*metadataprefix*) per especificar el format de metadades en el que es vol obtenir la resposta. A banda, permet introduir altres opcions com ara el rang de dates de les dades que es volen recuperar.
- *Listrecords*: ofereix una informació del conjunt de totes les dades i metadades del repositori.
- *Getrecord*: recupera un registre concret. Necessita dos arguments: identificador del registre sol·licitat (*identifier*) i especificació del format de metadades (*metadataprefix*).
- *Listsets*: informa dels *datasets* que conformen integren l'OAI-PMH. És el més utilitzat ja que, a banda de donar una visió global del repositori, es poden seleccionar els identificadors i els registres de cadascun dels conjunts de dades.

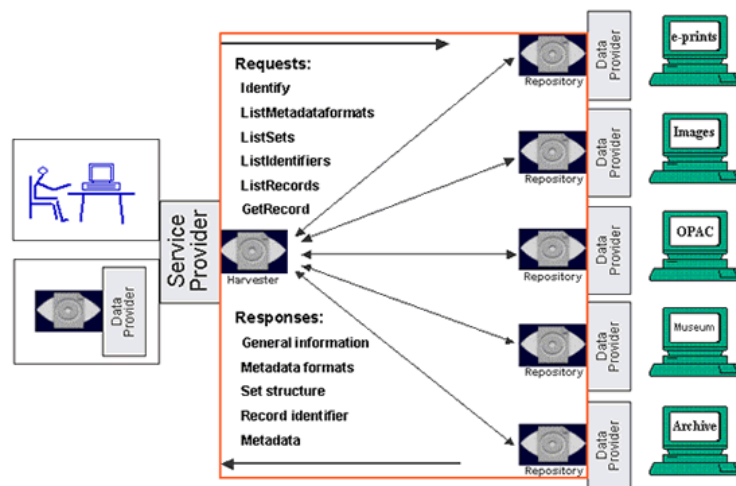


Figura 8. Visió de conjunt i model d'estructura de l'OAI-PMH. Font: [Open Archive Initiative](http://www.openarchives.org/PMH/)

Així, mitjançant aquestes peticions es poden filtrar i seleccionar l'enviament de determinades metadades segons determinats criteris com la data de creació dels registres o una determinada matèria o concepte. En resposta, el proveïdor de dades retorna un conjunt de registres en format xml codificat i estructurat.

3 L'Arxiu Nacional de Catalunya

A través del Decret 314/1980 es creava l'Arxiu Nacional de Catalunya, adscrit a la Direcció General de Patrimoni Cultural del Departament de Cultura i Mitjans de Comunicació de la Generalitat de Catalunya, i es definia com l'arxiu històric que havia d'aplegar, conservar, tractar, fer accessible i difondre la documentació històrica pública o privada generada a Catalunya per l'Administració i per altres institucions.



Figura 9. Actual seu de l'Arxiu Nacional de Catalunya. Foto: J. Play

Fins l'any 1995 va desenvolupar les seves tasques en un edifici de l'Eixample de Barcelona, moment en què es va produir el seu trasllat a la seva actual seu, en un edifici de nova planta amb les instal·lacions adients per allotjar-hi un arxiu, al municipi de Sant Cugat del Vallès, amb una superfície total de 12.625 m² i amb capacitat per custodiar-hi fins a prop més de 77.000 metres lineals de documentació.

A més d'assegurar la gestió, l'ús i la salvaguarda dels arxius del Govern i de l'Administració pública de Catalunya, l'Arxiu Nacional de Catalunya aplega, conserva i difon el patrimoni documental de Catalunya.

Com a arxiu general de l'Administració, ingressa i gestiona la documentació generada per l'acció política i administrativa del Govern, departaments, organismes i empreses de la Generalitat de Catalunya i és considerat l'arxiu final, on els arxius centrals han de dipositar la documentació de conservació permanent que generen els serveis centrals i els territorials de la circumscripció de Barcelona.

En la seva condició d'arxiu històric, recupera tota aquella altra documentació que té un relleu especial per al coneixement i la interpretació de la història col·lectiva de Catalunya.

3.1 Missió i funcions

Tal i com es recull en la seva carta de serveis la seva missió consisteix en:

1. Administrar, salvaguardar i preservar els arxius i registres del Govern i l'Administració pública de Catalunya, així com els d'individus, d'organitzacions i d'entitats catalanes per tal que les persones descobreixin, usin i aprenguin d'aquest patrimoni documental.
2. Afavorir la consolidació d'una societat lliure, oberta i democràtica en assegurar l'accés als documents, instruments essencials que garanteixen l'exercici dels drets dels ciutadans i el control de les actuacions dels poders públics.
3. Facilitar la comprensió històrica de la identitat nacional catalana i donar suport a la vida cultural i intel·lectual dels catalans.

A la llei 10/2001, d'arxius i documents de Catalunya, també es descriuen les seves funcions i activitats:

1. Rebre la documentació en fase semiactiva d'ús infreqüent (documentació que, un cop conclosa la tramitació ordinària, no és utilitzada d'una manera habitual per la unitat que l'ha produït) i de conservació permanent, i la documentació històrica dels òrgans centrals de l'Administració de la Generalitat i de les entitats i les empreses que en depenen, preservar-la i tenir-la a disposició de l'Administració i dels ciutadans.

2. Ingressar, conservar i difondre els fons i els documents privats que per llur valor testimonial i referencial concerneixin Catalunya i siguin d'una rellevància especial.
3. Promoure les activitats de difusió del patrimoni documental que custodia i establir, quan es consideri escaient, els acords que calgui amb altres institucions per fomentar-ne el tractament tècnic i la difusió.

3.2 Organització

L'Arxiu Nacional de Catalunya s'estructura en 8 àrees:

1. Àrea dels fons de l'administració: conserva i gestiona els testimonis documentals produïts per l'Administració pública a Catalunya, especialment durant el segle XX.
2. Àrea dels fons històrics: recupera i gestiona tota aquella documentació que té un relleu especial per al coneixement de la història col·lectiva de Catalunya.
3. Àrea dels fons d'imatges, gràfics i audiovisuals: dona el tractament específic als documents d'imatge i sonors en els seus diversos suports.
4. Àrea de reprografia i noves tecnologies: duu a terme els programes de reprografia dels fons, alhora que fa el seguiment de les innovacions tecnològiques en el camp de la informació i el tractament documental, i elabora, si s'escau, els programes pertinents per a la seva implementació a l'Arxiu.
5. Àrea de restauració: duu a terme les tasques de control, preservació, conservació i consolidació material dels documents que custodia l'Arxiu, alhora que impulsa programes de preservació documental en col·laboració amb d'altres institucions públiques i privades.
6. Àrea d'informació: proporciona els recursos d'informació i assessorament als usuaris i gestiona el servei de consulta als fons documental, bibliogràfics i hemerogràfics de l'Arxiu. També coordina el funcionament de la sala de consulta i de la biblioteca.
7. Àrea d'acció cultural: en la vessant de projecció, difusió i comunicació duu a terme totes aquelles accions de tipus cultural encaminades a donar a conèixer als diferents segments de la societat els fons documentals i els continguts de l'arxiu.
8. Servei didàctic: potencia entre el professorat i l'alumnat de secundària la didàctica de la història amb fons d'arxiu. Ho fa mitjançant la utilització dels fons documentals de l'arxiu com a font per a l'estudi i la recerca i elaborant recursos documentals i didàctics específics.

3.3 Serveis que ofereix

L'Arxiu Nacional de Catalunya ofereix els següents serveis:

1. Informació general i assessorament especialitzat sobre els fons documentals, bibliogràfics i hemerogràfics, els instruments de descripció, les condicions d'accés als fons i els serveis de l'arxiu.
2. Informació de referència sobre altres centres d'arxiu i fons documentals
3. Consulta pública i gratuïta dels fons documentals, bibliogràfics i hemerogràfics de l'arxiu i dels seus instruments de descripció. Aquests últims es poden consultar presencialment a la seu de l'Arxiu o a través del cercador que la institució posa a disposició dels seus usuaris a Internet. En cas que els documents objecte de la consulta estiguin digitalitzats i associats a les seves descripcions (i sempre i quan no hi hagi restriccions d'accés), aquests poden ser visualitzats en línia.
4. Reproducció de documents: aquest servei s'ofereix amb els límits, tarifes i formes de pagament establerts per la legislació i normativa vigents. El servei de reprografia permet obtenir fotocòpies, impressions d'imatges digitals, digitalitzacions, impressions de microfilms i duplicació de documents sonors i de vídeo, sempre i quan l'estat de conservació dels documents ho permeti i no hi hagi cap tipus de restricció per fer-ho.
5. En la seva condició d'Arxiu General de l'administració de Catalunya, presta servei als titulars dels fons públics facilitant-los l'accés a la informació i als documents que hi ha dipositats a l'Arxiu. Aquests usuaris institucionals poden accedir a la informació i als documents a través de tres modalitats de servei: la consulta, el préstec i la recuperació.
6. Vinculat amb el seu Servei Didàctic, l'Arxiu ofereix assessorament al professorat per a la preparació de materials didàctics; assessorament als alumnes que realitzen un treball de recerca; organització de visites de grups escolars i tallers didàctics.

3.4 Fons que s'hi custodien

L'Arxiu Nacional de Catalunya ingressa i custodia fons d'orígens molt diversos, tant pel que fa a la seva procedència, titularitat, règim jurídic, suport i format.

Així, l'Arxiu conserva prop més d'un miler de fons i col·leccions documentals que ocupen més d'una trentena de quilòmetres de prestatgeria. Pel que fa a arxius d'imatge i so, l'Arxiu aplega més de cinc milions d'imatges i prop de vint-i-cinc-mil registres sonors i audiovisuals. També es conserven a l'Arxiu uns cinquanta-cinc-mil cartells, dibuixos i gravats i uns cent-seixanta-mil mapes i plànols.

La classificació dels fons es fa d'acord amb el quadre d'organització general establert pel Departament de Cultura de la Generalitat de Catalunya i les seves descripcions s'adapten a

la norma ISAD (G) (Norma Internacional de Descripció Arxivística) i a la NODAC (Norma de Descripció Arxivística de Catalunya).

Entre els fons documentals custodiats per l'Arxiu es podrien destacar:

1. Fons de la Generalitat de Catalunya: s'hi pot trobar la documentació produïda per la Generalitat durant la Segona República (1931-1939); la produïda per la Generalitat a l'exili (1939-1977); i la documentació produïda per la Generalitat des del seu restabliment l'any 1977 fins a l'actualitat.
2. Fons de l'administració local: en aquest grup de fons s'inclou la documentació vinculada a la institució que es considera el precedent contemporani de la Generalitat de Catalunya. Així, hi trobem el fons de la Mancomunitat de Catalunya (1914-1925) i el fons de la Comissió Liquidadora de la Mancomunitat de Catalunya (1925-1932) durant la dictadura de Primo de Rivera per tal de gestionar la liquidació dels comptes de l'extingida Mancomunitat.
3. Fons de l'administració perifèrica de l'estat: formen part d'aquest grup els fons de les delegacions ministerials a l'àmbit provincial de Barcelona. Es tracta de la documentació generada a Catalunya durant el franquisme i transferida a la Generalitat amb els traspassos de competències.
4. Fons de l'administració reial i senyorial: l'Arxiu conserva més de 3.000 ples de la sèrie processos civils procedents de la Reial Audiència de Catalunya, des de finals del segle XVI fins al XIX.
5. Fons judicials: s'hi pot trobar documentació de diferents òrgans judicials radicats a Catalunya al llarg del segle XX.
6. Fons registrals: fons constituïts pels 2.507 llibres-registre dels Corredors Reials de Canvi de Barcelona (1780-1956).
7. Fons d'institucions: són fons d'organismes creats per una funció molt específica, de servei o d'interès públic, vinculats en major o menor grau a la Generalitat de Catalunya.
8. Fons religiosos: l'Arxiu aplega documentació d'institucions religioses com a resultat d'aquells períodes històrics en què aquestes patiren la dissolució del seu patrimoni.
9. Fons d'associacions i fundacions: inclouen arxius d'associacions i entitats, organitzacions sindicals, i de partit i associacions polítiques que han tingut importància en la història contemporània de Catalunya.
10. Fons comercials i d'empreses: actualment l'Arxiu conserva un centenar de fons d'empreses de diferents sectors, a partir del segle XVIII, entre els que destaquen el tèxtil, el metal·lúrgic, el comercial, el químic i l'editorial.

11. Fons patrimonials i familiars: destaquen per la seva importància els nobiliaris. Entre aquests cal destacar el llinatge Setnmenat, Marquesos Castellidosrius o el llinatge Moixó. El conjunt d'aquests fons apleguen una col·lecció de pergamins que arriba a les 7.000 unitats.
12. Fons personals: a l'Arxiu Nacional de Catalunya s'hi troba també un conjunt d'arxius i fons de personalitats de la societat catalana dels segles XIX i XX d'àmbits molt diversos: polítics (Francesc Macià, Ventura Gassol i Ramon Trias Fargas, entre d'altres), intel·lectuals (Carles Riba o Ferran Soldevila) o artistes (com Joaquim Renart). Entre els fons personals també s'hi pot trobar l'obra fotogràfica de diversos fotògrafs professionals i afeccionats des de principis de segle: Josep Gaspar, els Branguí, Gabriel Casas o Bert i Claret.
13. Col·leccions: l'Arxiu també aplega col·leccions factícies provinents de recuperacions, donacions i compres. Entre aquestes cal destacar la col·lecció d'auques de Pere Grañén i Raso o la col·lecció de cartells sobre prevenció de riscos d'accidents laborals. L'ANC també conserva col·leccions factícies de manuscrits, pergamins i impresos.

3.5 Activitats i publicacions

L'Arxiu Nacional de Catalunya també organitza activitats, conferències i exposicions. Entre aquestes, caldria destacar les jornades de portes obertes que organitza cada mes de juny coincidint amb la Setmana Internacional dels Arxius o les diferents exposicions que ha organitzat com la relativa a la commemoració del vint-i-cinquè aniversari de l'Arxiu o la que va servir per explicar el procés de confiscació que va patir Catalunya arran de la Guerra Civil, la història de la reivindicació del retorn d'aquesta documentació i la documentació retornada.

A banda, també edita publicacions que descriuen alguns dels fons més rellevants custodiats per l'Arxiu i disposa d'un butlletí, que es publica des de l'any 2002 cada quadrimestre, i que recull informació sobre els principals ingressos, els fons documentals més importants, notícies d'actualitat o activitats organitzades per l'Arxiu.

3.6 La pàgina web de l'ANC

L'Arxiu Nacional de Catalunya disposa d'una [pàgina web](#) amb informació bàsica de la institució, novetats, agenda d'activitats, normativa interna o enllaços als perfils de facebook i twitter de la institució. Aquesta s'estructura en sis blocs:

1. Coneix l'ANC: en aquest bloc s'hi recull informació relativa a la missió de l'Arxiu, com s'organitza, els projectes que du a terme i enllaços a la carta de serveis de la institució i a les seves memòries d'activitat.
2. Fons: secció que permet conèixer la riquesa i diversitat de fons documentals que aplega l'Arxiu.
3. Serveis: s'expliquen els serveis de consulta oberts a la ciutadania, els de reprografia de documents i els serveis específics per al sector educatiu i per a la pròpia administració.
4. Acció cultural: es recullen les activitats de projecció i difusió, les publicacions i els materials multimèdia que fa l'Arxiu o hi participa.
5. Gestió: és una secció dedicada al sector professional de l'arxivística i a programadors culturals interessants en les exposicions produïdes per l'arxiu.
6. Actualitat: és l'espai on es troben les notícies vinculades amb l'activitat desenvolupada per l'Arxiu i que s'han anat publicant a la seva pàgina web.
7. Horaris i contacte: aquí s'hi pot trobar informació relativa a la ubicació de l'Arxiu, contacte i horaris d'atenció a l'usuari.

L'actual pàgina web de l'Arxiu Nacional de Catalunya és fruit del redisseny del portal que es va produir l'any 2011, dotant-la d'una nova estructura conceptual i funcional i fent-la més àgil i flexible que la que disposava la institució fins aquell moment²⁵ (ARXIU NACIONAL DE CATALUNYA (Editorial), 2011).

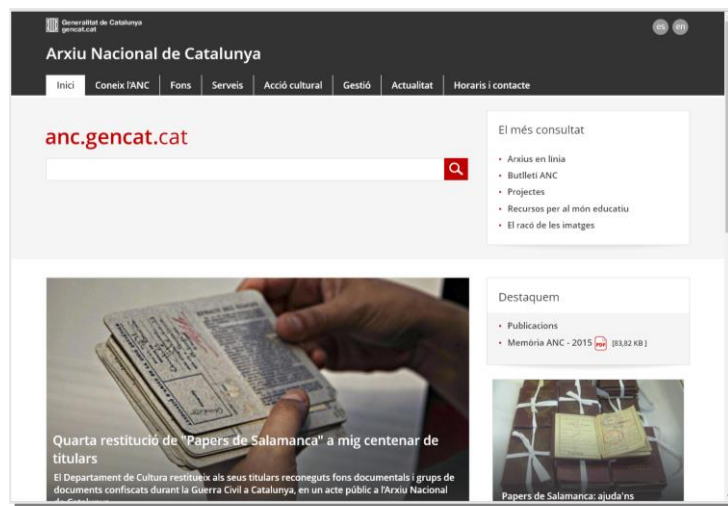


Figura 10. Vista de la pàgina inicial del portal web de [l'Arxiu Nacional de Catalunya](http://anc.gencat.cat)

²⁵ Des de l'any 2002 l'Arxiu disposa de pàgina web però, fins l'any 2011, el lloc web estava integrat dins del web del Departament de Cultura de la Generalitat de Catalunya (PETIT CIBIRAIN, 2003).

3.7 Els cercadors

A la mateixa pàgina web s'hi poden trobar enllaços als cercadors de l'Arxiu per tal que els usuaris puguin accedir a la descripció dels fons documentals i col·leccions que s'hi custodien. Els cercadors que s'ofereixen són els que es detallen a continuació:

1. CercANC: cercador operatiu a través d'Arxius en línia (que es tractarà específicament en el pròxim punt del treball)
2. ANC@ula: aquest cercador té finalitats educatives i facilita la cerca i la visualització de documents de l'Arxiu, alhora que es proposa contribuir a la utilització dels documents en la pràctica docent i en les recerques escolars.
3. BEG (Catàleg de Biblioteques Especialitzades de la Generalitat de Catalunya): és el catàleg conjunt format per 35 biblioteques i Centres de Documentació dels diferents departaments de la Generalitat. El catàleg de l'Arxiu inclou el fons bibliogràfic i hemerogràfic auxiliar i de reserva que són adients a les diferents tipologies dels fons documentals ingressats a l'arxiu. El catàleg de la biblioteca conté més de 48.000 registres i disposa de més de 200 capçaleres de publicacions periòdiques.

3.7.1 CercANC i Arxius en línia

El cercador de l'Arxiu Nacional de Catalunya és operatiu a través d'Arxius en línia, el cercador de fons i documents dels arxius del Departament de Cultura de la Generalitat de Catalunya.

Aquesta eina dona a conèixer els fons i documents que es custodien a l'Arxiu Nacional de Catalunya, a la Xarxa d'Arxius Comarcals, als Arxius Històrics Provincials, al Dipòsit d'Arxius de Cervera, als Arxius Històrics de l'Aran, Hostalric i Cardona i als Arxius Municipals de Castelló d'Empúries, Montornès del Vallès, Palafrugell i Sant Feliu de Guíxols.

El sistema permet cercar els documents independentment del centre on es troben, del fons al que pertanyen i del tipus de document de què es tracti. També permet el visionat de les imatges associades a les descripcions dels documents i ofereix la possibilitat de reservar la documentació en línia.

Arxius en línia permet diferents opcions de cerca:

1. Cerques simple i avançada: s'accedeix als documents que compleixen els paràmetres de la cerca.
2. Cerca per inventari: recupera la informació per fons documentals i, desplegant els diferents nivells del quadre de classificació, es pot accedir als documents concrets.

Des de la cerca simple es poden fer cerques genèriques tipus Google i, des de l'avançada, se'n poden fer de més acotades a partir d'una gran varietat de combinacions. Aquesta última està pensada per usuaris familiaritzats amb el món dels arxius.

Els resultats de la cerca permeten la visualització mitjançant una llista de documents o, sempre que els documents cercats estiguin digitalitzats i associats a la descripció, a través d'una galeria d'imatges.

El cercador presenta la informació del fons tenint en compte els principis de la descripció multinivell. El concepte sobre el qual s'organitza la cerca i la recuperació de la informació és el fons documental. Cada fons inclou la descripció dels documents que el formen, independentment de la seva naturalesa textual o no textual.

Val a dir que, si bé Arxius en línia està en funcionament des de finals de l'any 2009, des de l'any 2002 l'Arxiu Nacional de Catalunya posa a disposició dels seus usuaris la consulta dels fons documentals que custodia a través d'Internet (ARXIU NACIONAL DE CATALUNYA (Editorial), 2002). Fins aquell moment la pàgina web de l'ANC, operativa des de l'any 1999, era una pàgina molt senzilla i estàtica que tenia com a única finalitat presentar la institució als ciutadans. A partir de l'any 2002 l'usuari de l'ANC ja va poder consultar els instruments de descripció de l'arxiu des de la mateixa pàgina web de l'ANC que, com s'acaba de dir, es trobava dins la pàgina web del Departament de Cultura (a les descripcions de la documentació custodiada pels arxius comarcals s'hi accedia a través d'un altre cercador).

Per poder oferir la consulta en línia dels diferents instruments de descripció es va haver de fer, prèviament, una tasca d'unificació de totes les bases de dades existents en una única estructura que en possibilités la migració a una fitxa descriptiva estàndard igual per a tots els fons (CASTAN, A.; PETIT, C.; UTGÉS, M., 2009). Aquesta fitxa era la que estava definida en l'aplicació informàtica de l'Arxiu Nacional de Catalunya, el GANC.

La publicació, l'any 2007, de la NODAC (Norma de Descripció Arxivística de Catalunya) va fer que els arxius catalans comencessin a adaptar les seves descripcions a la norma i que el cercador de l'Arxiu Nacional quedés obsolet (ja que també s'havia d'adaptar a aquesta norma, integrant la descripció multinivell i la recuperació de la informació en el seu context documental).

Així, l'any 2008 es va constituir un equip de treball format per arxivers de l'ANC conjuntament amb personal de l'Àrea de Tecnologies de la Informació i les Comunicacions del Departament de Cultura, que van treballar en un nou cercador que es va posar en funcionament el mes d'octubre de l'any 2009 i que presentava la informació dels fons de manera més integrada i tenint en compte els principis de la descripció multinivell.

És també, i com s'acaba de citar més amunt, durant l'any 2009 quan es fan els primers intents per avançar cap a un cercador únic a través del qual es pugui tenir accés a totes les descripcions dels documents custodiats pels arxius històrics de la Generalitat de Catalunya. El mes de novembre de 2009 és quan es posa el nou cercador Arxius en línia, que substitueix els anteriors cercadors de l'ANC i de la XAC. Així, el nou cercador de l'ANC elaborat l'any 2009, s'integra, el mes de novembre d'aquell mateix any, al nou portal Arxius en Línia de la Subdirecció General d'Arxius i Gestió Documental.

Amb tot, malgrat que l'accés a les descripcions dels arxius històrics de la Generalitat de Catalunya es fa des d'un únic punt d'accés (Arxius en línia), en l'actualitat la duplicitat de programaris es manté i des dels arxius només són visibles les dades del propi centre.

3.8 El sistema informàtic de l'Arxiu Nacional de Catalunya

3.8.1 GANC (Gestor de l'Arxiu Nacional de Catalunya)

El sistema informàtic actual de l'Arxiu Nacional de Catalunya (com el de la Xarxa d'Arxius Comarcals) es va iniciar l'any 1993, amb la constitució del primer grup de treball format per tècnics d'ambdues institucions i del Servei d'Arxius (FERRER, J., MAGRINYÀ, A., SERNA, E., 2014). L'objectiu d'aquest grup era el de definir els requeriments bàsics que havia de tenir l'aplicació informàtica dels arxius històrics de la Generalitat de Catalunya. El grup va decidir desenvolupar, a partir d'uns requeriments comuns, un aplicatiu diferent per a cadascuna de les institucions: un per l'ANC (el GANC) i l'altre per la Xarxa d'Arxius Comarcals (el GAC).

Abans d'aquesta data l'Arxiu Nacional de Catalunya utilitzava diversos programes d'ofimàtica on es consignaven, com bonament es podia, els inventaris i catàlegs dels fons que es descrivien (BERNAL CERCÓS, A., 2009).

Entre els anys 1993 i 2000, doncs, el GANC es desenvolupa notablement. Es tracta d'un producte integral que gestiona tot el procés arxivístic, és a dir la gestió dels ingressos, l'organització dels fons (classificació, descripció i instal·lació) així com la gestió de l'espai. Al mateix temps, possibilita la consulta dels fons i dels inventaris i catàlegs des dels terminals que hi ha a disposició dels usuaris presencials a la sala de consulta de l'Arxiu o a Internet a través d'Arxius en Línia.

GANC també incorpora un mòdul per a la documentació no textual, permet associar imatges a les descripcions de forma individual o massiva, té la possibilitat de realitzar canvis massius per fer modificacions a un gran nombre de registres (sense haver-ho de fer de manera individual) i té un mòdul d'importació de dades externes que facilita l'automatització dels ingressos i transferències a l'arxiu procedents de l'administració.

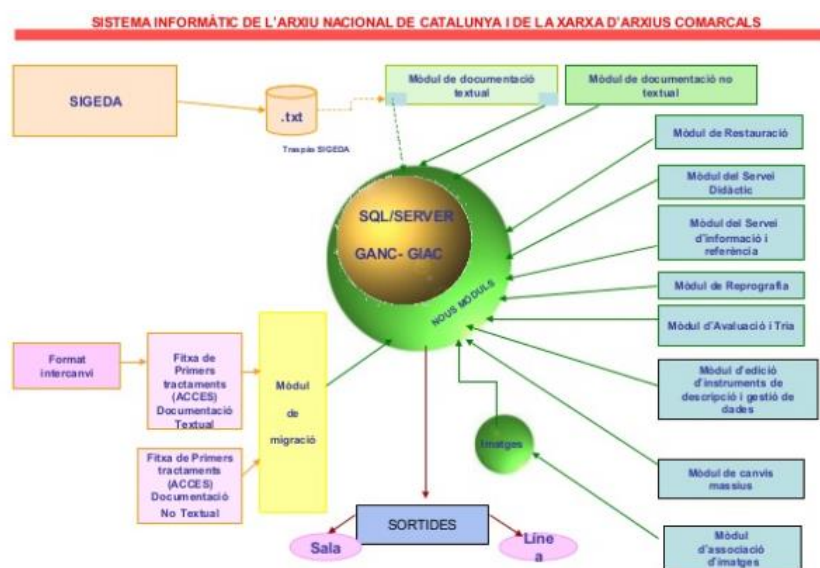


Figura 11. El sistema informàtic de l'ANC i de la Xarxa d'Arxius Comarcals. Font: Imma Navarro i Molleví.

Ponència presentada sota el títol [“El cas de l'Arxiu Nacional de Catalunya i els Arxius Comarcals. Una experiència profitosa”](#) a la II Jornada sobre la fotografia en Museus i Arxius celebrada el 30 de maig de 2013 al Museu Marítim de Barcelona

3.8.2 La fitxa de primers tractaments i el format d'intercanvi de dades

A l'Arxiu Nacional de Catalunya també es disposa d'una fitxa de primers tractaments (en format ACCESS)²⁶ que facilita als arxivers que hi treballen l'entrada de dades, l'extracció de diferents llistats, la possibilitat de realitzar comprovacions i correccions massives dels registres que hi ha a la base de dades i la migració de la informació continguda a aquesta fitxa de primers tractaments al GANC.

Aquesta base de dades disposa de 56 camps. Alguns d'aquests camps són instrumentals i poden acabar concatenats en d'altres camps com els de descripció o observacions. De tots aquests camps només en migren a GANC 34 i, en relació a la NODAC, entre els 56 camps es consigna informació dels elements obligatoris que dicta la norma: codi de referència, nivell de descripció, títol, dates, volum i suport, productor, autoria i dates.

Per migrar els registres que hi ha a la fitxa de primers tractaments a GANC, es fa des del mòdul de migració del mateix GANC, on hi ha la opció d'importació d'unitats de catalogació, i que és des del lloc on es pot seleccionar la base de dades que conté les unitats de catalogació que es volen migrar al GANC. Prèvia migració al GANC, es fan una sèrie de comprovacions perquè la informació es pugui integrar al sistema. És en el moment de la migració on es consigna la informació relativa a l'accessibilitat de cadascuna de les unitats

²⁶ Aquí es fa referència a la fitxa de primers tractaments per la documentació textual. Hi ha una altra fitxa de primers tractaments a l'ANC per la documentació no textual (amb unes altres característiques adaptades a la descripció d'aquests tipus de documents).

de descripció que hi ha en l'instrument a migrar: descripció consultable a sala (sí/no) i descripció consultable a Internet (sí/no).

Per altra banda, l'ANC també disposa d'un format d'intercanvi de dades que permet que els arxius centrals de la Generalitat de Catalunya que transfereixen la documentació que generen i que és de conservació permanent, puguin transferir també els registres informàtics que descriuen els documents transferits.

Durant els primers anys de l'existència de l'ANC i fins l'any 2000, els registres transferits es lliuraven en bases de dades de formats i estructures diverses (BERNAL CERCÓS, A., 2008). A partir del mes de juliol de 2000 l'ANC va definir les característiques del format d'intercanvi actual, vàlid per a fitxers bdf i mdb. Es tracta d'una estructura estàndard de dades mínimes, compatible amb les normes de descripció, a la qual cal traspasar els registres de descripció que s'han de transferir a l'arxiu històric des dels diferents sistemes informàtics de gestió de processos documentals que s'utilitzen als departaments i organismes de la Generalitat.

Les bases de dades rebudes per l'ANC es revisen, sistematitzen, milloren i preparen per migrar la informació al GANC, per tal que l'usuari hi pugui accedir des del mòdul de sala de consulta o des d'Arxius en Línia si les condicions d'accés a les descripcions així ho permeten.

Per acabar val la pena remarcar que malgrat l'Arxiu Nacional de Catalunya no té facultats per establir normes, marcar pautes o guiar projectes, l'activitat que s'hi desenvolupa ha marcat la pauta de molts projectes del Sistema d'Arxius de Catalunya (BERNAL CERCÓS, A., 2014). L'exemple del GANC és un exemple paradigmàtic del que s'acaba de citar, ja que la seva experiència fou exportada a altres arxius (cas dels comarcals amb l'aplicació amb la que treballen actualment), així com també el format d'intercanvi de dades, el mòdul de sala de consulta o el cercador a Internet.

3.9 Algunes dades de l'Arxiu Nacional de Catalunya

A continuació es detallen algunes dades que mostren l'evolució de l'activitat de l'Arxiu Nacional de Catalunya entre els anys 2011 i 2015. Les dades s'han extret de la memòria anual publicada pel Departament de Cultura de la Generalitat de Catalunya l'any 2015.

3.9.1 Volum total de documentació custodiada a l'Arxiu Nacional de Catalunya

L'Arxiu Nacional de Catalunya custodia prop més de 1.100 fons documentals que ocupen més de trenta quilòmetres de prestatgeria.

	2011	2012	2013	2014	2015
Volum físic					
Nombre acumulat de metres lineals dels fons	32.301	32.929	33.951	34.857	32.266
Nombre de metres lineals de documents ingressats	2.046	628	1.203	906	1.409
Fotografia	3.004.718	3.025.869	3.558.421	4.512.043	5.320.963
Sonor i audiovisual	22.567	23.849	24.355	24.898	24.975
Cartells, dibuixos, gravats i altres	53.408	53.239	53.374	53.594	55.411
Mapes i plànols	159.509	159.259	159.530	159.532	159.981
Dipòsit digital (repositori segur)					
Nombre de fitxers	3.867.968	5.292.044	5.749.909	6.332.429	6.540.743
Terabytes (total)	27,64	46,44	47,7	57,9	60,1

Taula 2. Volum de documentació custodiada a l'ANC (2011-2015). Font: [Memòria del Departament de Cultura de la Generalitat de Catalunya de l'any 2015. Síntesi 1: Dades d'activitat](#)

3.9.2 Volum total de documentació descrita a l'Arxiu Nacional de Catalunya

No tota la documentació que es custodia a l'ANC està descrita a nivell d'unitat documental composta o simple. Sí que en la majoria de casos està disponible la descripció a nivell de fons, fet que permet que l'usuari de l'Arxiu pugui disposar de la informació bàsica per saber què pot trobar en qualsevol dels fons documentals de l'ANC.

Malgrat això, el volum total d'unitats documentals descrites pels tècnics que treballen a l'Arxiu no és gens menyspreable²⁷.

	2013	2014	2015
Unitats de descripció textuals	1.090.024	1.238.799	1.375.280
Unitats de descripció no textuals (imatge i so)	289.866	299.675	308.080
Total unitats de descripció descrites	1.379.890	1.538.474	1.683.360
Documents textuals digitalitzats	675.163	780.188	803.162
Documents no textuals (imatge i so) digitalitzats	248.759	270.632	294.333
Total documents digitalitzats	923.922	1.050.820	1.097.495

Taula 3. Descripció dels fons de l'ANC. Unitats documentals. 2013-2015. Font: [Memòria del Departament de Cultura de la Generalitat de Catalunya de l'any 2015. Síntesi 1: Dades d'activitat](#)

Amb data 4 de maig de 2016 aquestes dades havien variat lleugerament: el total d'unitats de descripció descrites era 1.916.947 i el total de documents digitalitzats 1.394.366.

²⁷ Les dades que es presenten a la taula 3 són les que estan disponibles al mòdul principal de GANC. No són les mateixes, per raons d'accés, que estan disponibles a Arxius en Línia (informació que es mostra en la taula 4).

3.9.3 Volum total de documents consultables a Internet (Arxius en Línia)

No totes les descripcions dels documents de l'ANC estan disponibles a Internet. Moltes vegades ens podem trobar amb descripcions que estan en el sistema però que només són accessibles internament (només poden accedir-hi els treballadors de la institució). Això es deu al fet que algunes d'aquestes descripcions poden tenir dades de caràcter personal que no són accessibles a la xarxa per raons d'accés.

A continuació es mostra una taula amb les descripcions que l'Arxiu Nacional posa a l'abast dels seus usuaris a través de la xarxa.

	2013	2014	2015
Unitats de descripció textuals	282.801	372.294	452.181
Unitats de descripció no textuals (imatge i so)	195.887	217.421	226.010
Total unitats de descripció descrites	478.688	589.715	678.191
Documents textuals digitalitzats	372.442	465.530	494.413
Documents no textuals (imatge i so) digitalitzats	170.715	201.731	219.669
Total documents digitalitzats	543.157	667.261	714.082

Taula 4. Unitats de descripció disponibles a la xarxa. 2013-2015. Font: [Memòria del Departament de Cultura de la Generalitat de Catalunya de l'any 2015. Síntesi 1: Dades d'activitat](#)

Amb data 4 de maig de 2016 aquestes dades havien variat lleugerament: el total d'unitats de descripció disponibles a la xarxa era 724.561 i el total de documents digitalitzats disponibles a la xarxa era 928.293.

3.9.4 Usuaris de l'ANC i consultes de les descripcions

A continuació es mostra l'evolució dels usuaris de l'Arxiu Nacional de Catalunya i el total de consultes que es fan a l'Arxiu (ja sigui presencialment o a través dels cercadors en línia).

	2011	2012	2013	2014	2015
Investigadors/usuaris	3.222	2.912	2.999	2.974	3.005
Titulars de fons	327	304	244	303	267
Usuaris remots (visites web)	-	-	29.497	59.923	95.876
Total	3.549*	3.216*	32.740	63.200	99.148

Taula 5. Usuaris de l'ANC entre els anys 2011 i 2015. Font: [Memòria del Departament de Cultura de la Generalitat de Catalunya de l'any 2015. Síntesi 1: Dades d'activitat](#)

*Les dades relatives als usuaris remots (visites web) no estan disponibles pels anys 2011 i 2012 ja que només es disposa d'aquesta informació a partir de l'any 2013

	2011	2012	2013	2014	2015
Sala de consulta	8.651	8.729	7.796	13.121	12.111
Servei d'informació i referència	4.568	4.120	5.019	5.076	5.140
Internet (Arxius en Línia)	-	-	55.946	86.861	138.977
Total	13.219*	12.849*	68.761	105.058	156.228

Taula 6. Consulta i comunicació fons de l'ANC. Unitats documentals 2011-2015. Font: Font: [Memòria del Departament de Cultura de la Generalitat de Catalunya de l'any 2015. Síntesi 1: Dades d'activitat](#)

*Les dades relatives a les consultes realitzades a través d'Arxius en Línia no estan disponibles pels anys 2011 i 2012 ja que només es disposa d'aquesta informació a partir de l'any 2013

Les dades acabades d'exposar en aquestes taules mostren el gran volum d'informació estructurada (instruments de descripció arxivístics) que una institució com l'Arxiu Nacional de Catalunya posa a disposició dels seus usuaris (quasi 2 milions d'unitats de descripció de les quals 725.000 estan disponibles a Internet).

Al mateix temps, es pot apreciar com:

- És molt més gran el nombre d'usuaris que consulten aquestes descripcions a través de la xarxa sense desplaçar-se a la seu de l'ANC.
- També és més alt el percentatge de les consultes en línia respecte les consultes presencials (havent-se incrementat notablement entre els anys 2014 i 2015).
- I com les consultes en línia han incrementat exponencialment al llarg dels anys (quasi triplicant-se l'any 2015 respecte l'any 2013).

	2013	2014	2015
Sala de consulta	7.796 (12,2%)	13.121 (13,1%)	12.111 (8%)
Internet (Arxius en Línia)	55.946 (87,8%)	86.861 (86,9%)	138.977 (92%)
Total	63.742 (100%)	99.982 (100%)	151.088 (100%)

Taula 7. Consultes d'unitats documentals a l'ANC. 2013-2015. Font: [Memòria del Departament de Cultura de la Generalitat de Catalunya de l'any 2015. Síntesi 1: Dades d'activitat](#)

4 Proposta d'obertura de dades a l'Arxiu Nacional de Catalunya

Fins aquí s'han exposat els conceptes de dades obertes, les raons i les avantatges d'obrir dades, quina és la legislació relativa a la reutilització de la informació del sector públic a nivell europeu i estatal i s'han analitzat diferents iniciatives d'arxius d'arreu del món en l'obertura de dades. També s'ha explicitat la missió, visió i valors de l'Arxiu Nacional de Catalunya, els serveis que ofereix i s'han exposat dades dels volums de documentació que

s'hi custodia, el total de descripcions que es posen a l'abast dels usuaris i el nombre de consultes que realitzen aquests últims.

Tenint en compte doncs, d'una banda, la importància i les avantatges que suposa pel ciutadà disposar de dades estructurades i en format obert i, de l'altra, el paper que juga l'ANC dins del sistema arxivístic català (o almenys el que se li podria reclamar com a arxiu nacional del país), en aquest últim apartat del treball es presentarà l'anàlisi i el disseny d'un projecte d'obertura de dades per l'Arxiu Nacional de Catalunya que permeti millorar les prestacions que ofereix l'ANC als seus usuaris.

Així, s'exposarà com es mostra actualment la informació dels instruments de descripció de l'ANC a Internet (Arxius en Línia), s'analitzaran les possibles mancances del cercador, es mostrarà, a partir de peticions reals fetes a l'Arxiu, com disposar de dades en obert és una demanda a l'alça, i s'intentarà demostrar com el fet d'oferir els instruments de descripció de l'ANC en formats reutilitzables no suposaria una inversió massa elevada ni en recursos humans ni econòmics, ja que dita inversió es realitza en el moment de descriure els documents i elaborar els inventaris i catàlegs que l'Arxiu ja ofereix a la ciutadania, i no en el moment de canviar el format amb el que es presenten. Finalment, es mostraran els resultats obtinguts d'una prova de concepte feta sobre la plataforma *open source* CKAN de 9 conjunts de dades que es proposen com a primer pas d'obertura de dades a l'Arxiu Nacional de Catalunya, motivant les raons per les quals s'han triat aquests 9 *datasets* per fer la prova de concepte.

4.1 Com es mostra la informació a Arxius en Línia

En el punt 3.7.1 del present treball s'ha descrit el cercador Arxius en Línia i les possibilitats que ofereix. Tot seguit es mostrarà com es presenten els resultats de les cerques que fan els usuaris i què es permet fer amb els resultats obtinguts.

Per cercar a arxius en línia hi ha tres opcions: cerca simple, cerca avançada i cerca per inventari.

1. Cerca per inventari (figura 12): permet buscar informació sobre fons documentals. Es pot seleccionar si el fons que es busca conté documents textuais o no textuais, quin contingut (temàtica) es vol, dates extremes que interessin i, en cas de voler-se buscar la informació en un centre arxivístic concret, seleccionar-lo a través d'un desplegable. Els resultats que s'obtenen són llistats de fons documentals on en qualsevol camp de la seva descripció hi ha la informació consignada prèviament en els criteris de cerca (figura 13). Els resultats apareixen ordenats per centre arxivístic. Selecciónant el fons que més interressi s'accedeix a la seva descripció segons la

Norma de Descripció Arxivística de Catalunya (NODAC) i, en cas que hi hagi unitats documentals descrites associades al fons seleccionat, a la part esquerra de la pantalla es pot navegar pel quadre de classificació associat al fons (figura 14). Seleccionant els diferents nivells del quadre es pot accedir als títols dels documents associats a cada entrada.

Figura 12. Pantalla de cerca per inventari a Arxius en Línia

Fons de l'Arxiu Nacional de Catalunya	Tipus de document	Dates de creació	Dates d'agregació
ANC1-288	FRANCESC IGLEAS I PIERNAU - FLORA IGLEAS I ALSINA - ENRIC ADROHER I PASQUAL (GIRONELLA)	Dates de creació: 1920 - 1991 (predomina 1938 - 1989)	
ANC1-963	FRANCESC RIBÓ I ROVIRA - MARIANA HONCLÓS I CHARLES	Dates de creació: 1927 - 1974 (predomina 1937 - 1955)	
ANC1-810	FRANCESC SERRADO I CANAL	Dates de creació: 1905 - 1978	
ANC1-630	FREDERIC ESCOFFET	Dates de creació: 1936 - 1985	
ANC1-89	FREDERIC BAHOLA	Dates de creació: 1940 - 1984	
ANC1-1028	FUNDACIÓ ARTUR MARTORELL	Dates de creació: 1969 - 2012	
ANC1-511	GENERALITAT DE CATALUNYA (EXILI)	Dates de creació: [1914-1918], 1932-1940, 1945, 1965-1977	
ANC1-1	GENERALITAT DE CATALUNYA (SEGONA REPÚBLICA)	Dates de creació: 1939-1940, 1945, 1945-1977	Dates d'agregació: Predomina 1931-1939. Inclou, excepcionalment, documentació anterior (1756, 1826, 1862/1869, 1908, 1911 1917-1930) i posterior (1940 i 1945). Hi ha índexs d'actes del Consell Executiu fets al 1983 per OIBA.
ANC1-216	GOVERN CIVIL DE BARCELONA	Dates de creació i d'agregació: 1931-1939.	
ANC1-494	GRANJA ESCOLA D'AGRICULTURA	Dates de creació i d'agregació: 1950-1979.	

Figura 13. Resultats obtinguts a partir d'una cerca per inventari



Figura 14. Descripció de fons a Arxius en Línia amb possibilitat de navegar pel quadre de classificació

2. Cerca simple (figura 15): aquesta cerca actua contra els diferents camps descriptius de les unitats documentals descrites de qualsevol fons. Es pot cercar per contingut, dates, decidir si els documents es presenten llistats (tots els documents) o en galeria (es mostren els documents que tenen la seva imatge associada) i, en cas de voler-se buscar la informació en un centre arxivístic concret, seleccionar-lo a través d'un desplegable. El resultat de la cerca és un llistat de fons documentals, ordenats per centre arxivístic (figura 16), on clicant sobre cadascun d'ells es mostren els títols de les unitats documentals que compleixen els paràmetres de la cerca (figura 17).



Figura 15. Pantalla de cerca simple a Arxius en Línia

Generalitat de Catalunya
gencat.cat

Departament de Cultura

El Departament
Els nostres serveis
Tot cultura
Paràmetres de cerca
Contingut: segona república

La Cultura ets tu

inici > Els nostres serveis > Base de dades > Arxius i Gestió Documental > Arxius en línia > Cerca simple

Cerca simple Cerca avançada Cerca per inventari

Llista de fons de la cerca simple

S'han trobat 40 fons documentals i 302 unitats documentals

Pàgina 1/2

Fons de l'Arxiu Nacional de Catalunya

Fons i col·leccions documentals que l'arxiu custodia.

		Documents textuals	Documents no textuals	Unitats amb imatges digitalitzades
ANCI-64	BERT I CLARET (FOTOPERIODISTES)	0	1	1
ANCI-262	CAMBRA OFICIAL DEL LLEBRE DE BARCELONA	2	0	0
ANCI-236	CENTRE PENITENCIARI D'HOMES DE BARCELONA (PRESÓ MODEL)	1	0	0
ANCI-555	CENTRO ESPAÑOL DE MOSCÚ (AGE)	1	50	51
ANCI-132	COL·LECCIÓ FOTOGRÀFICA D'AGUSTÍ DURAN I SANPERE	0	1	1
ANCI-372	DISPENSARI DR. LLUÍS SAYÉ (DISPENSARI ANTITUBERCULÓS)	1	0	0
ANCI-892	EDUARD PONS PRADES	1	0	0
ANCI-149	FEDERACIÓ NACIONAL D'ESTUDIANTS DE CATALUNYA (FNEC)	2	0	0
ANCI-1121	FRANCISC XAVIER PERAN I BRUGUÉS	0	1	1
ANCI-5	GABRIEL CASAS I GALDABARDES	0	17	17
ANCI-511	GENERALITAT DE CATALUNYA (EXILI)	2	0	1
ANCI-1	GENERALITAT DE CATALUNYA (SEGONA REPÚBLICA)	6	1	4
ANCI-726	JOAN FONTANILLAS GINABREDA	0	8	8
ANCI-783	JOAQUIM MARIA DE NADAL I FERRER	1	0	0
ANCI-244	JOSEP BENET	2	0	0

Figura 16. Resultats obtinguts a partir d'una cerca simple

Generalitat de Catalunya
gencat.cat

Departament de Cultura

El Departament
Els nostres serveis
Tot cultura
Paràmetres de cerca
Contingut: segona república

La Cultura ets tu

inici > Els nostres serveis > Base de dades > Arxius i Gestió Documental > Arxius en línia > Cerca simple

Cerca simple Cerca avançada Cerca per inventari

Arxiu Nacional de Catalunya
FONS ANCI-1 / GENERALITAT DE CATALUNYA (SEGONA REPÚBLICA)

Tipus de fons: Generalitat de Catalunya

D'aquest fons s'ha trobat 1 document

PRESIDÈNCIA (inclou Departament de la Presidència) > ADMINISTRACIÓ GENERAL > Legislació i assumptes jurídics

Pàgina 1/1

Codi	Títol	Núm. Expedient(s)/Signatura	Data(es)
ANCI-1-1-7501	Expedient sobre conflictes de competències amb el Ministeri d'Hisenda i Economia.		[19.07.1937 / 27.01.1938]

v2.1.2 | Avís legal | Accessibilitat | Sobre gencat | © Generalitat de Catalunya

Figura 17. Visualització del títol d'unitat documental a Arxius en Línia

3. Cerca avançada (figura 18): permet fer cerques més acotades a partir d'una gran varietat de combinacions. Així, es pot buscar per document textual o no textual, es poden fer cerques contra el camp títol o descripció, cercar dates concretes o per unitat productora o sèrie documental de determinat fons. Aquest tipus de cerca està pensada per usuaris familiaritzats amb el món dels arxius i la visualització que s'obté dels resultats obtinguts és la mateixa que la cerca simple.

Figura 18. Pantalla de cerca avançada a Arxius en Línia

En relació a les unitats documentals, Arxius en Línia, per defecte, mostra només el títol d'aquesta. Hi ha una, “visualització extensa”, que permet a l'usuari veure part de la descripció associada a la unitat documental. En cas que hi estigui interessat, clicant sobre el títol de la unitat descriptiva s'obre una pantalla on es pot visualitzar la seva descripció completa (figura 19).

Figura 19. Visualització de la descripció d'una unitat de catalogació a Arxius en Línia

En les descripcions de les unitats documentals que l'Arxiu Nacional posa a l'abast del ciutadà a Arxius en Línia hi consten els elements que, segons la NODAC, es consideren indispensables en una descripció arxivística: codi de referència, nivell de descripció, títol, data (es) i volum i suport. En molts casos també hi consten elements superiors als

considerats essencials però aquests depenen de la naturalesa de la unitat de descripció i, sobretot, dels recursos humans disponibles per haver pogut ampliar la descripció abans de penjar-la a Internet. Aquests últims serien elements com ara l'abast i contingut, la llengua i l'escriptura dels documents o les característiques físiques i requeriments tècnics, entre d'altres.

Des de la pantalla de descripció de les unitats de catalogació, l'usuari pot accedir al formulari de reserva de documentació (en format word). Aquest formulari l'usuari se l'ha de descarregar, consignar els diferents camps per sol·licitar la documentació objecte del seu interès i fer-lo arribar per correu electrònic a l'Arxiu Nacional.

4.1.1 Algunes mancances d'Arxius en Línia pels usuaris del segle XXI

Com es pot apreciar, Arxius en Línia assoleix els objectius bàsics de la descripció arxivística en línia: facilitar l'accés als documents des de qualsevol lloc i en qualsevol moment. A més, permet, en molts casos, consultar les imatges associades a les descripcions, cosa que fa que l'usuari no s'hagi de desplaçar a l'arxiu i pugui consultar la documentació des de casa o des del seu lloc de treball.

Ara bé, tenint en compte el context en què ens trobem i per tot el que s'ha exposat en aquest treball, l'aplicatiu té una sèrie de mancances per l'usuari. Les més rellevants potser serien:

- Espai personal de l'usuari que possibiliti desar cerques que s'han fet amb anterioritat, seleccionar unitats de descripció, imprimir-ne la descripció, agrupar-les per temes o per rangs de dates extremes i, fins i tot, poder fer comandes a reprografia d'aquells documents que estiguin digitalitzats i tinguin la seva imatge associada.
- Possibilitat de veure en pantalla unitats de descripció que pertanyin a diferents fons i poder-les seleccionar en funció del fons, les dates extremes, el seu nivell de descripció o l'arxiu on es localitzin.
- Possibilitat d'ordenar els resultats de les cerques per rellevància, identificador o títol.
- Opció de compartir les cerques realitzades a través de twitter, facebook, linkedin o gmail.
- Possibilitat d'exportar els resultats obtinguts en formats oberts (csv, json, pdf o xml) i descarregar-los al propi maquinari per poder-los reutilitzar.

Aquest darrer punt està estretament relacionat amb la temàtica del present treball. Qualsevol usuari hauria de poder descarregar-se les descripcions que hi ha disponibles a Arxius en Línia per reutilitzar-les i poder millorar les seves pròpies bases de dades, creuar-les amb altres dades de què pugui disposar i estalviar temps de manera considerable.

S'ha de tenir en compte que, tal i com es mostra actualment la informació a Arxius en Línia, l'usuari, si vol treballar amb alguna de les dades que consten a les descripcions, ha de tornar a picar la informació que s'hi mostra o haver de copiar i enganxar els resultats obtinguts de les cerques.

4.2 Disposar de formats reutilitzables: una demanda a l'alça

En els darrers mesos l'Arxiu Nacional de Catalunya ha rebut més d'una sol·licitud en aquesta línia. Tot seguit es descriuen algunes de les peticions que s'han rebut en aquest sentit.

El mes d'octubre de l'any 2015 l'Arxiu Nacional publicava, en format pdf, la [Guia de la sèrie documental Procediments Judicials Militars \(Sumaríssims\) 1939-1980](#), resultat de 10 anys de tasques de tractament i descripció arxivística dels expedients judicials militars instruïts a Catalunya durant la dictadura franquista i que es conserven a la seu del Tribunal Militar Territorial Tercer de Barcelona²⁸. A la guia, a banda de quedar explicitat l'abast i contingut de la sèrie segons les normes internacionals de descripció arxivística, s'hi incloïen diversos resums estadístics (per edat, sexe, lloc de naixement i lloc de residència, entre d'altres) sobre les persones processades.

La publicació va tenir una molt bona acollida per part dels investigadors i ciutadans en general i molts mitjans de comunicació se'n van fer ressò.

Un d'aquests mitjans, el "Punt Avui", es va posar en contacte amb l'Arxiu Nacional de Catalunya per demanar si podia disposar de les estadístiques que es presentaven a la guia en un format més reutilitzable per publicar-les al seu diari i desenvolupar altres productes a partir de les dades. La persona que es va posar en contacte amb l'Arxiu volia ordenar les dades de manera diferent a la que apareixien a la guia i volia elaborar un mapa de persones represaliades a Catalunya per demarcacions. Amb el format amb el que estava la guia (pdf) li era difícil reutilitzar la informació sense perdre temps.

²⁸ En dit tractament hi col·laboraren el mateix Tribunal Militar, titular de la documentació, el Memorial Democràtic de la Generalitat de Catalunya, organisme encarregat de proporcionar els recursos humans i materials, i l'Arxiu Nacional de Catalunya, responsable de coordinar el tractament i aportar-hi la direcció tècnica. El tractament va consistir en instal·lar correctament els expedients per a la seva adequada conservació i informatitzar en una base de dades Access tots els noms de les persones processades que constaven en un fitxer manual ubicat a la Secretaria del Tribunal Militar Territorial Tercer, per tal d'agilitar-ne les consultes. A banda de bolcar les dades de les persones processades a la base de dades, es va ampliar la informació que hi havia a les fitxes manuals, amb la consulta directa dels procediments on apareixien aquestes persones. La fitxa on es bolcaven les dades era la fitxa de primers tractaments de l'ANC adaptada a les peculiaritats de la sèrie a tractar.

Des de l'Arxiu Nacional es va respondre afirmativament a la sol·licitud que feia aquest mitjà i se li va passar la informació que apareixia a la guia en format xls²⁹. A continuació es mostra un exemple de la informació que va poder elaborar el diari a partir de les dades facilitades.

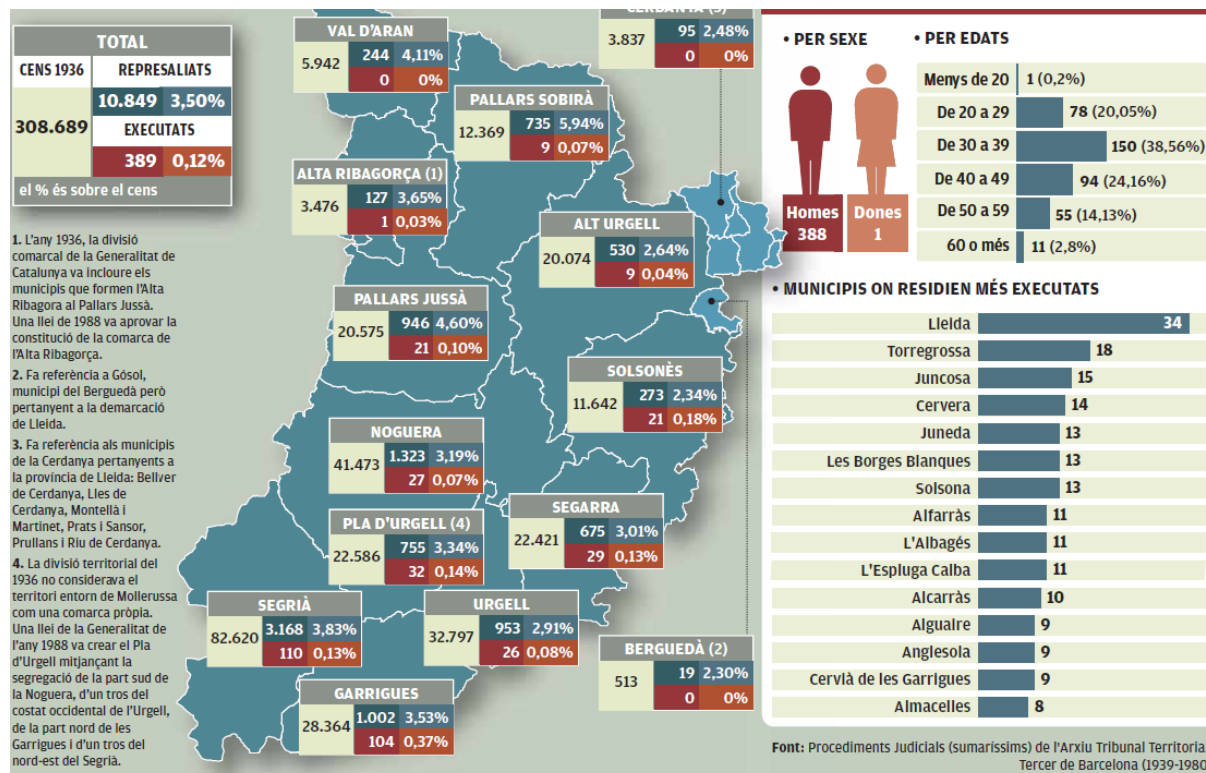


Figura 20. Mapa de persones represaliades i executades a les comarques de Lleida elaborat pel "Punt Avui" a partir de les dades de la Guia Procediments Judicials Militars (sumarissims) del Tribunal Militar Territorial Tercer de Barcelona. Edició de Lleida del Punt Avui del 8 de novembre de 2015.

Per aquelles mateixes dates Òmnium Cultural organitzava un acte en memòria de les víctimes del franquisme. Els organitzadors de l'esdeveniment es van adreçar a l'Arxiu Nacional de Catalunya per sol·licitar el fet de poder disposar en un format més reutilitzable la relació de persones executades a Catalunya a partir d'un procediment judicial militar que apareixia a la guia. L'Arxiu Nacional va respondre afirmativament a la sol·licitud i el 20 de novembre, en el transcurs de l'acte davant de la Presó Model de Barcelona, es va poder veure un immens plafó amb el nom de les persones executades pel franquisme a Catalunya que apareixien a la guia.

Finalment, aquest mateix 2016, un altre usuari de l'Arxiu que estava realitzant un estudi sobre concessions mineres, va demanar poder disposar d'un arxiu en format excel amb les

²⁹ Les dades publicades a la guia van sortir de l'explotació de la informació continguda en la base de dades (en format Access) resultant del tractament. Es van elaborar estadístiques a partir de consultes que anaven contra la taula principal on hi havia la informació i es van incrustar en un document Word que es va passar a pdf. Per passar-les al diari el "Punt Avui" es van passar les dades de l'Access a l'Excel.

dades dels expedients de mines del fons 1-641 de l'ANC³⁰, de la zona del Berguedà. Necessitava les mateixes dades que apareixien a Arxius en Línia (número d'expedient, municipi, nom de la mina o data de petició de la concessió, entre d'altres) i es volia estalviar la feina de picar-les de nou. L'Arxiu Nacional de Catalunya també va donar una resposta afirmativa a aquesta demanda de l'usuari i se li va facilitar un excel amb la informació que sol·licitava. Per fer-ho es van agafar les dades del fons 1-641 de GANC Llistats³¹ i es van passar a excel.

4.3 Definició d'un projecte d'Open Data per l'Arxiu Nacional de Catalunya

Amb tot el que s'ha exposat fins aquí es fa força evident que l'Arxiu Nacional de Catalunya hauria de poder oferir als seus usuaris conjunts de dades en formats reutilitzables.

Els punts anteriors han servit per contextualitzar la metodologia i les eines informàtiques que els arxivers de l'ANC utilitzen per descriure els fons i documents que estan dipositats a la institució amb la finalitat de posar les seves descripcions a l'abast de qualsevol ciutadà.

Com tot seguit es mostrarà, els instruments que s'utilitzen possibiliten que la conversió de la informació que hi ha a GANC i a Arxius en Línia en formats reutilitzables sigui relativament senzilla i ràpida.

En aquest darrer apartat del treball, doncs, es definirà un projecte d'obertura de dades per l'ANC, proposant, com a prova de concepte, l'obertura de 9 instruments de descripció que estan disponibles actualment al portal Arxius en Línia.

També es plantejarà, amb la voluntat que l'ANC es presenti davant la societat com una institució transparent i generadora de confiança, la possibilitat d'oferir altra tipus d'informació en format obert, com ara el total de consultes que rep l'ANC, el global d'unitats descrites que estan en línia o el volum de documentació custodiada per la institució.

Finalment es proposaran possibles línies a seguir de cara al futur en relació a l'obertura de dades i a la transparència.

³⁰ [Fons 1-641](#) Prefectura del Districte Miner de Barcelona.

³¹ GANC Llistats és una eina interna de l'Arxiu Nacional de Catalunya que va contra la informació que hi ha a GANC (l'aplicació informàtica de l'Arxiu Nacional de Catalunya). És una base de dades en format Access i amb diverses taules que permet extreure llistats i fer filtres amb les dades que hi ha a GANC.

4.3.1 Quines dades s'utilitzaran per fer la prova de concepte

Per fer la prova de concepte es proposa treballar amb 9 *datasets* que es corresponen amb les unitats de descripció de 9 fons gestionats per l'Àrea dels Fons de l'Administració de l'Arxiu Nacional de Catalunya sobre els que més consultes es fan a Arxius en Línia³².

S'han seleccionat aquests 9 instruments de descripció per dur a terme la prova de concepte degut al fet que, tal i com es mostra en la taula que segueix a continuació, han estat dels més consultats a Internet des que es tenen estadístiques de les consultes que fan els usuaris a Arxius en Línia (any 2012).

Fons	Consultes al fons	Nombre consultes a unitats textuais	Nombre consultes a imatges associades a u. text.	Nombre consultes a unitats no textuais	Nombre consultes a imatges associades a u. no text.
ANC1-42 - BRANGULÍ (FOTÒGRAFS)	35303 (6,74%)	0 (0,00%)	0 (0,00%)	197064 (36,47%)	178300 (38,37%)
ANC1-1 - GENERALITAT DE CATALUNYA (SEGONA REPÚBLICA)	33551 (6,40%)	28642 (14,58%)	56291 (34,08%)	22561 (4,18%)	20620 (4,44%)
ANC1-715 - JUNTA DE MUSEUS DE CATALUNYA	21092 (4,03%)	29315 (14,93%)	60013 (36,33%)	19 (0,00%)	13 (0,00%)
ANC1-5 - GABRIEL CASAS I GALO BARDES	17949 (3,43%)	0 (0,00%)	0 (0,00%)	67905 (12,57%)	61889 (13,32%)
ANC1-585 - JOSEP MARIA SAGARRA I PLANA	11140 (2,13%)	0 (0,00%)	0 (0,00%)	36060 (6,67%)	35744 (7,69%)
ANC1-402 - ANTONIO ESPLUGAS	11086 (2,12%)	0 (0,00%)	0 (0,00%)	19908 (3,68%)	11319 (2,44%)
ANC1-818 - PRESIDENT FRANCESC MACIÀ (documentació)	8432 (1,61%)	10235 (5,21%)	12053 (7,30%)	0 (0,00%)	0 (0,00%)
ANC1-64 - BERT I CLARET (FOTOPERIODISTES)	8305 (1,58%)	0 (0,00%)	0 (0,00%)	20253 (3,75%)	16899 (3,64%)
ANC1-398 - REIAL AUDIÈNCIA DE CATALUNYA	7815 (1,49%)	6754 (3,44%)	6606 (4,00%)	0 (0,00%)	0 (0,00%)
ANC1-1051 - BALTASAR PORCEL	7615 (1,45%)	3506 (1,79%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-167 - LLINATGE SENTMENAT, MARQUESOS DE	7541 (1,44%)	7053 (3,59%)	4936 (2,99%)	0 (0,00%)	0 (0,00%)
ANC1-610 - MINYONS ESCOLTES DE CATALUNYA	7448 (1,42%)	81 (0,04%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-264 - FRANCESC MACIÀ	7138 (1,36%)	1083 (0,55%)	0 (0,00%)	6048 (1,12%)	4305 (0,93%)
ANC1-633 - CONSELL NACIONAL DE LA JOVENTUT DE	7098 (1,35%)	2770 (1,41%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-555 - CENTRO ESPAÑOL DE MOSCÚ (AGE)	6260 (1,19%)	915 (0,47%)	231 (0,14%)	5907 (1,09%)	4782 (1,03%)
ANC1-255 - EUGENI D'ORS	6255 (1,19%)	0 (0,00%)	0 (0,00%)	5501 (1,02%)	3707 (0,80%)
ANC1-667 - JOAN TRIADÚ	6092 (1,16%)	3135 (1,60%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-651 - LLUÍS COMPANYYS	5788 (1,10%)	425 (0,22%)	394 (0,24%)	5010 (0,93%)	4373 (0,94%)
ANC1-960 - ARXIU DEL PALAU-REQUESENS	5667 (1,08%)	3485 (1,77%)	1115 (0,68%)	0 (0,00%)	0 (0,00%)
ANC1-211 - FUERZAS ELÉCTRICAS DE CATALUÑA, S.A	5442 (1,04%)	0 (0,00%)	0 (0,00%)	26617 (4,93%)	24452 (5,26%)
ANC1-57 - INSTITUT ESCOLA	5056 (0,96%)	525 (0,27%)	175 (0,11%)	4684 (0,87%)	3861 (0,83%)
ANC1-230 - PARTIT SOCIALISTA UNIFICAT DE CATALUNYA	4951 (0,94%)	3003 (1,53%)	885 (0,54%)	0 (0,00%)	0 (0,00%)
ANC1-737 - JOSEP PUIG I CADA FALCH	4851 (0,93%)	1753 (0,89%)	0 (0,00%)	1398 (0,26%)	3513 (0,76%)
ANC1-23 - JOSEP GASPAR I SERRA	4504 (0,86%)	0 (0,00%)	0 (0,00%)	8065 (1,49%)	6740 (1,45%)
ANC1-172 - LLINATGE D'ARQUER	4385 (0,84%)	5044 (2,57%)	3591 (2,17%)	0 (0,00%)	0 (0,00%)
ANC1-138 - COMPAÑÍA GENERAL DE TABACOS DE FILIPINAS,	4324 (0,83%)	0 (0,00%)	0 (0,00%)	4841 (0,90%)	3950 (0,85%)
ANC1-97 - CARLES SOLDEVILA	4319 (0,82%)	2012 (1,02%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-511 - GENERALITAT DE CATALUNYA (EXILI)	4284 (0,82%)	2770 (1,41%)	2398 (1,45%)	0 (0,00%)	0 (0,00%)
ANC1-253 - ALEXANDRE GALÍ	4132 (0,79%)	1402 (0,71%)	0 (0,00%)	898 (0,17%)	405 (0,09%)
ANC1-319 - FUERZAS HIDROELÉCTRICAS DEL SEGRE, S.A.	4036 (0,77%)	0 (0,00%)	0 (0,00%)	5294 (0,98%)	4248 (0,91%)
ANC1-757 - FERROCARRILS DE LA GENERALITAT DE	3901 (0,74%)	0 (0,00%)	0 (0,00%)	26772 (4,95%)	26006 (5,60%)
ANC1-886 - LLEI 21/2005 DE RESTITUCIÓ A LA GENERALITAT	3856 (0,74%)	1778 (0,91%)	115 (0,07%)	0 (0,00%)	0 (0,00%)
ANC1-367 - PAU CASALS	3742 (0,71%)	1289 (0,66%)	0 (0,00%)	877 (0,16%)	0 (0,00%)
ANC1-169 - LLINATGE ORIOLA-CORTADA, COMTES DE LA VALL	3740 (0,71%)	3191 (1,62%)	2289 (1,39%)	0 (0,00%)	0 (0,00%)
ANC1-547 - AGÈNCIA CATALANA DE L'AIGUA	3739 (0,71%)	2270 (1,16%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-26 - CARLES RIBA - CLEMENTINA ARDERIU	3347 (0,64%)	1238 (0,63%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-261 - CENTRE DE LA PROPIETAT INTEL·LECTUAL	3283 (0,63%)	2170 (1,10%)	1250 (0,76%)	0 (0,00%)	0 (0,00%)
ANC1-803 - MOVIMENT DE Joves CRISTIANS DE POBLES DE	2772 (0,53%)	44 (0,02%)	0 (0,00%)	0 (0,00%)	0 (0,00%)
ANC1-566 - COL·LECCIÓ NOTIZARIO DEL CORPO TRUPPE	2741 (0,52%)	2453 (1,25%)	1840 (1,11%)	0 (0,00%)	0 (0,00%)
ANC1-718 - COL·LECCIÓ DE CARTELLS DE JOVENTUT	2661 (0,51%)	0 (0,00%)	0 (0,00%)	1925 (0,36%)	1229 (0,26%)
ANC1-800 - ANTONI ROSAL I GRELLON	2633 (0,50%)	0 (0,00%)	0 (0,00%)	3461 (0,64%)	2765 (0,60%)
ANC1-471 - TRIBUNAL REGIONAL DE RESPONSABILITATS	2438 (0,47%)	1116 (0,57%)	928 (0,56%)	0 (0,00%)	0 (0,00%)

Taula 8. Consultes en línia realitzades sobre els fons de l'ANC. 2012-2016. Es ressalten els fons gestionats per l'Àrea dels Fons de l'Administració amb els que es proposa fer la prova de concepte. Font: ANC

³² A l'ANC es disposa d'una aplicació interna que permet extreure estadístiques d'Arxius en Línia i saber quins són els fons dels que més consultes en línia es realitzen.

Amb tot, es proposa “obrir”, en un primer moment, els següents fons:

1. **ANC1-1 – Generalitat de Catalunya (Segona República):** catàleg del fons generat pel Consell Executiu i els òrgans centrals dels departaments de la Generalitat de Catalunya durant la segona república (1931-1939). Estan disponibles en línia 11.351 unitats de catalogació.
2. **ANC1-715 – Junta de Museus de Catalunya:** catàleg del fons que descriu la documentació generada per la Junta de Museus de Catalunya entre els anys 1891 i 1985. Estan disponibles en línia 6.979 unitats de catalogació.
3. **ANC1-818 – President Francesc Macià (documentació institucional):** catàleg del fons generat pel gabinet del President Francesc Macià entre els anys 1931 i 1933. Estan disponibles en línia 12.592 unitats de catalogació.
4. **ANC1-398 – Reial Audiència de Catalunya:** catàleg del fons generat per la Reial Audiència de Catalunya entre els anys 1534 i 1849. Estan disponibles en línia 5.794 unitats de catalogació.
5. **ANC1-633 – Consell Nacional de la Joventut de Catalunya:** catàleg del fons generat pel Consell Nacional de la Joventut de Catalunya entre els anys 1979 i 1999. Estan disponibles en línia 3.534 unitats de catalogació.
6. **ANC1-57 – Institut Escola:** catàleg del fons generat per l'Institut Escola entre els anys 1931 i 1939. Estan disponibles en línia 1.124 unitats de catalogació.
7. **ANC1-511 – Generalitat de Catalunya (Exili):** catàleg del fons generat per la Generalitat de Catalunya a l'exili. Estan disponibles en línia 443 unitats de catalogació.
8. **ANC1-886 – Llei 21/2005 de restitució a la Generalitat de Catalunya:** catàleg de la documentació que fou requisada per les tropes franquistes a particulars, sindicats, partits polítics i altres entitats de Catalunya al finalitzar la guerra civil espanyola i que fou retornada a la Generalitat en virtut de la Llei 21/2005 de restitució a la Generalitat de Catalunya. Estan disponibles en línia 19.165 unitats de catalogació.
9. **ANC1-647 – Agència Catalana de l'Aigua:** catàleg del fons generat per l'Agència Catalana de l'Aigua entre els anys 1845 i 1999. Estan disponibles en línia 5.452 unitats de catalogació.

4.3.2 La legislació permet obrir aquestes dades?

Tal i com s'ha exposat més amunt, els arxivers de l'ANC descriuen la documentació textual a partir de la fitxa de primers tractaments, en format Access, i es migra a GANC a través del mòdul principal del programari. La fitxa de primers tractaments disposa de dos camps per indicar el tipus d'accessibilitat de cada unitat de descripció:

- Sala (Sí/No).
- Internet (Sí/No).

Així, ens podem trobar amb unitats de catalogació la descripció de les quals:

- Només està disponible internament: només els treballadors de l'ANC poden accedir a la seva descripció.
- Només està disponible a sala de consulta: només els usuaris presencials de l'Arxiu poden consultar-ne la seva descripció.
- Està disponible a sala de consulta i a internet: tant els usuaris presencials com els no presencials poden consultar-ne la seva descripció.

Els conjunts de dades que es proposen utilitzar per fer la prova de concepte d'obertura de dades (i que, en un futur, es farien extensibles a la resta de fons documentals de l'ANC amb aquestes mateixes característiques), estan disponibles a Arxius en Línia, amb la qual cosa la seva descripció és lliure i accessible per qualsevol ciutadà.

Amb tot, no s'està oferint informació que atempti ni contra la intimitat ni altres drets privats legítims de les persones, ni que perjudiqui aspectes com, entre d'altres, la seguretat pública, la investigació, o la sanció d'infraccions penals, administratives o disciplinàries, aspectes previstos a la Llei 10/2001, de 13 de juliol, d'arxius i gestió de documents i a la Llei 19/2014, de 29 de desembre, de transparència, accés a la informació pública i bon govern, i que en limitarien el seu accés. A la vegada, tampoc s'entra en col·lisió amb cap dret de propietat intel·lectual que en pugui restringir la seva obertura, reutilització i redistribució.

4.3.3 Termes d'ús i llicències de les dades que es proposen oferir

Tenint en compte el que s'acaba d'exposar, les dades relatives als instruments de descripció que l'Arxiu Nacional de Catalunya oferiria en obert no haurien de tenir cap tipus de restricció i les persones que en fessin ús podrien fer-ne el que volguessin, ja que l'objectiu de la seva obertura seria poder-les reutilitzar en qualsevol àmbit, ja sigui la recerca, la difusió o, fins i tot, la seva explotació econòmica. Amb tot, les dades estarien lliures de patents, copyright, i no estarien subjectes a drets de privacitat, seguretat o privilegis que puguin estar reglades per altres normes.

En general doncs, se n'hauria de permetre la seva reproducció, distribució, comunicació pública i transformació per fer obres derivades, per a tothom i sense cap tipus de limitació temporal, i sempre que no es contradigués amb la llicència o avís que pogués tenir algun dels instruments de descripció en obert oferts i que seria la que prevaldria.

Malgrat tot, sí que s'haurien de seguir unes condicions mínimes per reutilitzar les dades:

- Que el contingut de la informació no sigui alterat.

- Que no es desnaturalitzi el sentit de la informació.
- Que se citi la font.
- Que s'esmenti la data de l'última actualització.

Finalment, es podria instar a les persones que utilitzessin les dades amb qualsevol de les finalitats citades, a posar en coneixement de l'ANC els projectes derivats de l'ús dels conjunts de dades, per tal que se'n pogués fer difusió i s'incentivés el seu ús per part de nous actors.

4.3.4 El formats en què es proposa oferir les dades

Per tal que les dades que s'oferirien en obert per part de l'Arxiu Nacional de Catalunya complissin els principis de l'Open Data, aquestes haurien d'estar en formats no propietaris. És a dir, i tal i com s'ha explicat al llarg d'aquest treball, no poden dependre d'una empresa o d'una eina propietària d'una empresa.

Per tant, els formats en què s'oferien les dades en obert de l'Arxiu Nacional de Catalunya serien *xlsx*, *csv* i *xml* (tots ells explicats en el punt 2.1.2 del present treball).

4.3.5 Plataforma sobre la que es farà la prova de concepte

La plataforma seleccionada per fer la prova de concepte amb els 9 conjunts de dades és CKAN. Tal i com s'ha explicat en aquest treball, CKAN és una plataforma *open source* utilitzada per nombrosos governs, organitzacions i comunitats arreu del món per desenvolupar els seus portals de dades obertes.

Els passos seguits per fer-ho són els que es descriuen a continuació:

- S'han analitzat els camps que es posen a disposició de l'usuari a Arxius en Línia de cada unitat de catalogació. Aquests camps són:
 - Nom de l'arxiu
 - Codi i nom del fons
 - Tipus de fons
 - Descripció del nivell de procedència
 - Descripció del nivell de classificació
 - Codi de referència de la unitat de catalogació
 - Nivell de descripció
 - Títol de la unitat de catalogació
 - Cronologia
 - Descripció de la unitat de catalogació
 - Observacions

- S'ha fet una consulta a GANC_Llistats per cadascun dels fons amb els que es volia fer la prova a CKAN per treure'n un llistat en format Access amb els mateixos camps que visualitza l'usuari a Arxius en Línia³³.
- S'han exportat les dades de cadascuna d'aquestes bases de dades a xlsx i a xml.
- Els fitxers xlsx s'han passat a csv.
- S'han pujat els conjunts de dades resultants a la demo de CKAN dins la organització "Prova TFM". Per pujar les dades a CKAN és molt senzill. S'ha d'entrar a la plataforma com a administrador i dins la organització creada (en aquest cas se l'ha anomenat "Prova TFM") prémer la icona "Afegir conjunt de dades". Quan es fa, s'obre una pantalla on es pot consignar el títol del conjunt de dades (seria el nom de l'instrument de descripció), la seva descripció i altres metadades relacionades amb el conjunt de dades que es puja: etiquetes, llicència sota la que es posen les dades, enllaç al recurs i nom i correu electrònic de l'autor o persona que manté el *dataset*, entre d'altres. En el moment de pujar les dades, i malgrat es pugui consignar aquesta informació de manera manual, la plataforma reconeix automàticament el format dels fitxers.

4.3.6 Resultats de la prova de concepte realitzada sobre CKAN

Tot seguit es mostren els resultats obtinguts amb la prova de concepte.

La pantalla principal a CKAN per l'organització seria com la que es mostra a continuació:

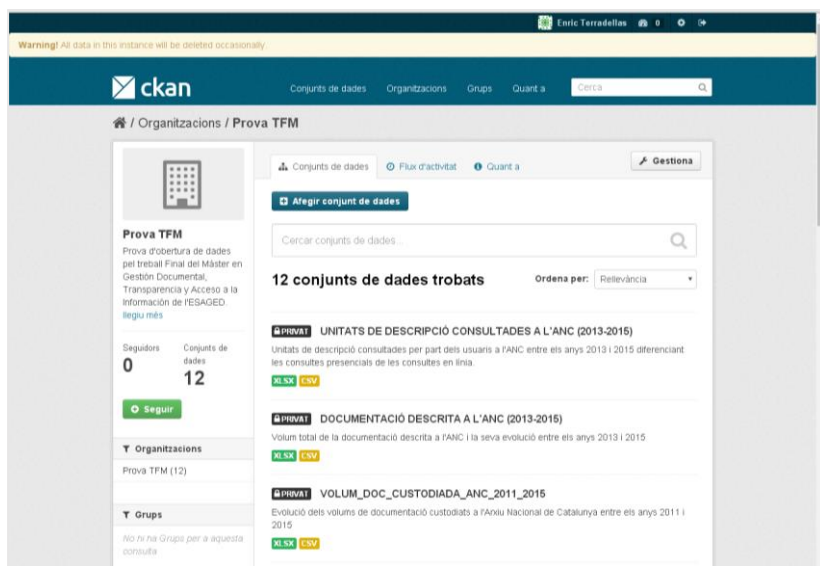


Figura 21. Visualització de la pantalla inicial per la organització Prova_TFM a CKAN

³³ Aquesta consulta va contra totes les unitats de catalogació que hi ha a GANC amb els filtres següents: número de fons (es selecciona el que interessa) i unitats de catalogació accessibles a Arxius en Línia (que la condició sigui que sí que són accessibles a Internet).

Des de la pantalla inicial es pot seleccionar qualsevol dels conjunts de dades pujats a la plataforma i, en cas de ser l'administrador i com s'ha dit abans, agregar conjunts de dades.

Com a usuari també es poden fer filtres de selecció dels conjunts de dades d'una organització a partir d'etiquetes, formats o tipus de llicències dels conjunts de dades. Aquesta selecció es pot fer a través d'un menú que apareix a la part esquerra de la pantalla principal per cada organització.

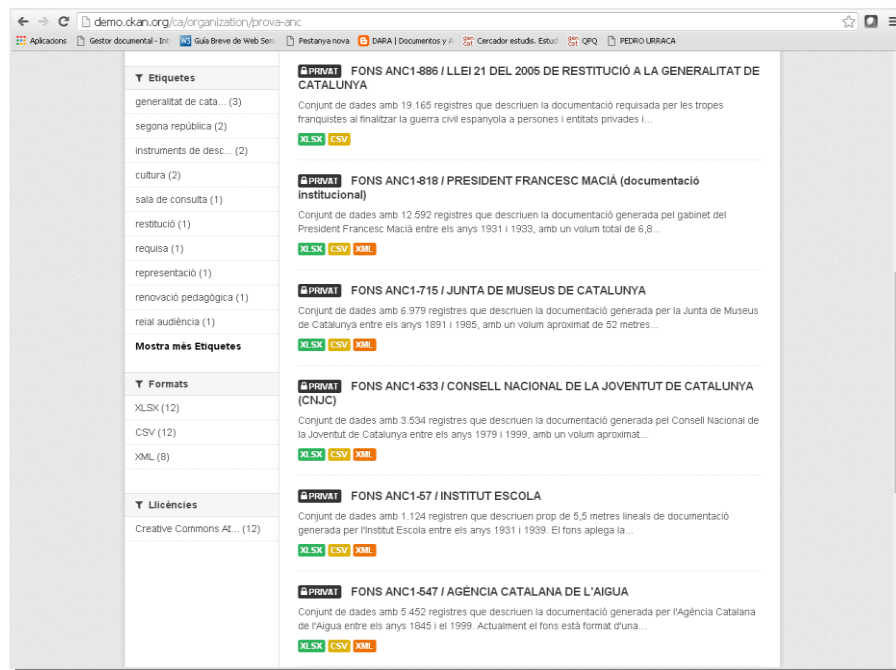


Figura 22. Visualització de la pantalla inicial per la organització Prova_TFM a CKAN. A la part esquerra es pot visualitzar la possibilitat de seleccionar conjunts de dades que duguin associades determinades etiquetes, segons els formats o llicències d'ús de les dades

En el moment de seleccionar qualsevol dels conjunts de dades que es mostren a la pantalla principal per cada organització, es pot visionar la descripció vinculada a aquest fons (s'ha fet constar en aquest camp la informació que apareix a la fitxa del fons a Arxius en Línia), accedir als diferents formats del conjunt de dades que hi ha disponibles i veure les dades de creació i última actualització del *dataset*, i la persona responsable de pujar-lo i el seu mantenidor.

Com a exemple, es mostra la informació que es veuria en pantalla en cas de seleccionar el Fons 1-398 / Reial Audiència de Catalunya.

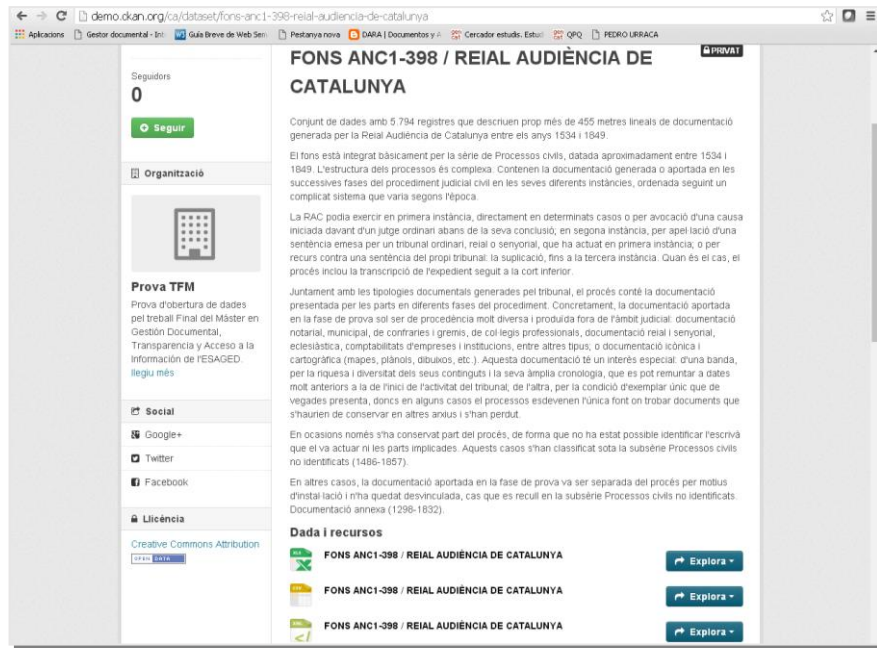


Figura 23. Visualització de la informació en el moment de seleccionar qualsevol conjunt de dades (en aquest cas Fons ANC 1-398 / Reial Audiència de Catalunya)

Per descarregar qualsevol dels conjunts de dades disponibles, només s'ha de seleccionar el format que més interressi a l'apartat "Dada i recursos". En fer-ho s'obre una pantalla on consta: una url des de la que, prement sobre ella, es pot descarregar el conjunt de dades a la màquina de l'usuari; un resum del conjunt de dades; la font del recurs; i una visualització estructurada de les dades que formen el conjunt.



Figura 24. Visualització del conjunt de dades per ser descarregat a CKAN

En cas que el responsable de pujar les dades a la plataforma hagi elaborat una visualització prèvia del recurs (com l'elaboració d'un gràfic o d'un mapa), es podria accedir a aquesta vista de les dades. Aquest últim aspecte és molt interessant ja que permet a les organitzacions que hi publiquen dades mostrar què es pot fer amb aquestes dades.

En relació a aquest últim punt, a banda de pujar-hi les dades en obert dels instruments de descripció dels 9 fons documentals esmentats, s'ha fet una prova per demostrar les possibilitats de visualització de la informació que ofereix CKAN. Així, en els fons ANC 1-1, ANC 1-398, ANC 1-547, ANC 1-715 i ANC 1-886 s'han geolocalitzat les unitats de descripció on estava informat el topònim.

En el moment de descriure la documentació, els arxivers de l'ANC poden personalitzar la fitxa de primers tractaments, afegint-hi camps on s'hi pot consignar informació que després es concatenarà al camp DESCRIPCIÓ o OBSERVACIONS. En funció del fons i dels recursos humans disponibles en el moment de descriure les unitats de descripció que conformen el fons, es pot crear un camp on s'hi consigna el topònim que pugui aparèixer al títol de les unitats documentals descrites o entre els documents que la formen. La informació que hi ha consignada a aquest camp es concatena, quan es migra a GANC, al camp "DESCRIPCIÓ" (darrera l'expressió "TOPÒNIMS:") juntament amb informació relativa al volum o a l'abast i el contingut. La informació dels topònims està normalitzada i sistematitzada i, a grans trets, es consigna sempre de la mateix manera:

- Pels topònims catalans: topònim seguit de la comarca i la província
- Pels topònims de la resta de l'estat: topònim seguit de la província
- Per la resta de topònims: topònim seguit del país

Per fer la prova de geolocalitzar unitats de descripció a CKAN s'ha aïllat (a partir de llenguatge Access) el topònim que constava al camp "DESCRIPCIÓ" de les unitats de catalogació dels fons acabats de citar i s'ha consignat massivament per cada topònim la seva latitud i longitud (en dos camps que s'han afegit a la base de dades extreta de GANC_LLISTATS)³⁴

A continuació es mostra una captura de pantalla de CKAN amb la visualització de les unitats de descripció del Fons 1-715 – Junta de Museus de Catalunya situades en un mapa:

³⁴ Les latituds i longituds de tots els municipis espanyols s'han extret de la pàgina web <http://www.businessintelligence.info/varios/longitud-latitud-pueblos-espana.html>.

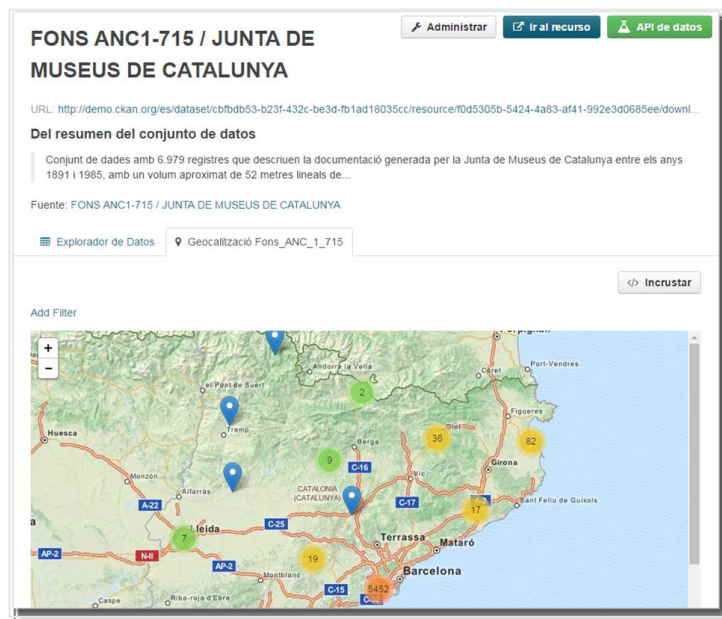


Figura 25. Visualització de la geocalització de les unitats de descripció del Fons 1-715 – Junta de Museus de Catalunya a partir de les possibilitats que ofereix la plataforma *open source* CKAN

Per últim, s'ha fet una darrera prova amb CKAN no relacionada amb l'obertura d'instruments de descripció de l'ANC: s'ha pujat en format obert indicadors d'activitat de l'Arxiu Nacional de Catalunya:

- Unitats de descripció consultades a l'ANC entre els anys 2013 i 2015.
- Total de documentació descrita a l'ANC entre els anys 2013 i 2015
- Volum de documentació custodiada a l'ANC entre els anys 2011 i 2015.

Com s'ha vist al llarg d'aquest treball, hi ha institucions arxivístiques que, a banda d'obrir els seus catàlegs i inventaris, estan aportant als portals de dades obertes dels seus respectius països informació com ara el nivell de satisfacció dels seus usuaris, el número de documents que digitalitzen cada any o l'evolució de les consultes que reben.

Tant l'Arxiu Nacional de Catalunya, com la Subdirecció General d'Arxius, com el Departament de Cultura de la Generalitat publiquen a les seves pàgines webs les seves memòries d'activitat anual. Aquestes memòries estan en format pdf, fet que fa que moltes de les taules que s'hi publiquin siguin difícilment reutilitzables.

A l'apartat 3.9 d'aquest treball s'han mostrat una sèrie de taules que apareixen publicades a la memòria del Departament de Cultura de la Generalitat de Catalunya de l'any 2015 relatives a l'activitat que es du a terme a l'Arxiu Nacional de Catalunya. Malgrat que la memòria es publica sota una llicència *creative commons*, es fa constar que la informació continguda a la memòria es podrà copiar i redistribuir en qualsevol mitjà i format però sense

fer-ne un ús comercial ni remesclant, transformant o creant algun producte a partir del material que conté la memòria.

Com a institucions de caràcter públic i tenint en compte, d'una banda, el que s'ha exposat en aquest treball, i de l'altra, el context en què ens trobem, aquesta informació hauria de publicar-se en formats oberts per tal que qualsevol ciutadà la pogués avaluar millor i, en cas d'interessar-l'hi, reutilitzar.

Aquesta última ha estat la raó per la que s'ha volgut fer la prova a CKAN de publicar indicadors d'activitat de l'ANC en formats *open data* (xlsx i csv).

A continuació es mostren les captures de pantalla de com es veurien a CKAN alguns dels indicadors de l'activitat de l'ANC que s'acaben de citar i les gràfiques que qualsevol ciutadà podria fer ràpidament en cas que estigués interessat en aquesta informació i se la descarregués des de CKAN:

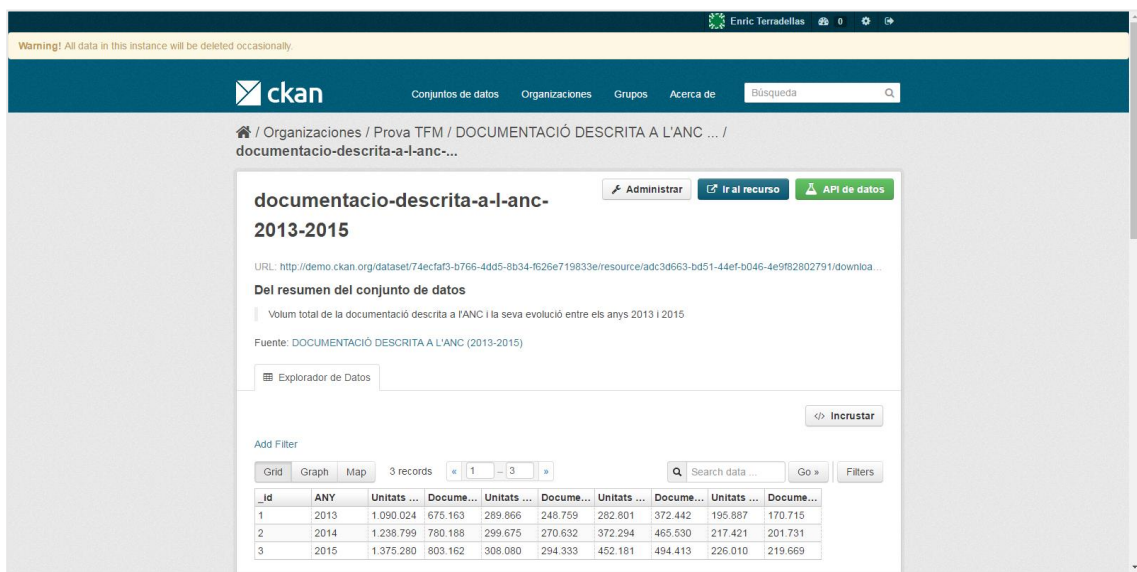


Figura 26. Visualització de la informació relativa a la documentació descrita a l'ANC entre els anys 2013 i 2015 a la plataforma open source CKAN en format xlsx i susceptible de ser descarregada per qualsevol ciutadà

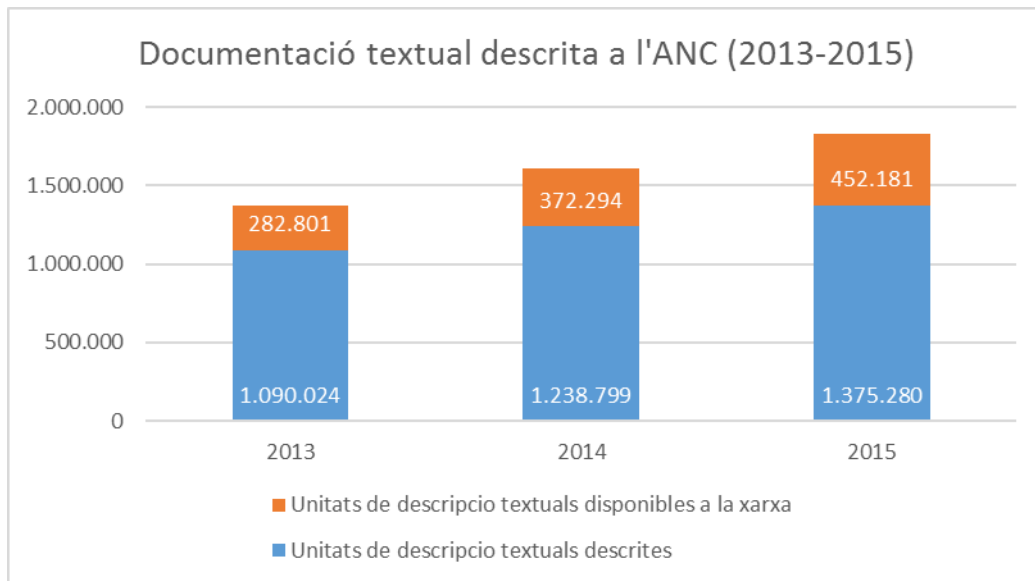


Figura 27. Gràfic que es podria realitzar fàcilment agafant les dades disponibles a CKAN relatives a la documentació textual descrita a l'ANC

A partir de la prova de concepte feta a CKAN s'ha pogut comprovar les funcionalitats d'aquesta plataforma *open source*:

- Permet pujar-hi dades en diferents formats oberts i de manera fàcil i intuïtiva.
- A partir de la consignació de determinades metadades per part del publicador de les dades, l'usuari final pot navegar a través d'etiquetes, formats i llicències d'ús de qualsevol dels conjunts de dades disponibles.
- L'usuari es pot descarregar a la pròpia màquina qualsevol dels conjunts de dades de qualsevol organització que hi publiqui dades en obert.
- La plataforma permet filtrar dades o elaborar gràfics a partir de les mateixes, per avaluar així les possibilitats d'explotació que ofereixen les dades publicades sense haver-les de descarregar prèviament.
- Permet a les organitzacions que hi publiquen dades mostrar les possibilitats d'aquestes últimes a partir de visualitzacions prèviament elaborades (com mapes o gràfics).

4.3.7 Valoració de la prova de concepte i ús de CKAN per obrir dades a l'ANC

Com s'ha anat exposant en l'apartat anterior del treball, CKAN és una plataforma *open source* que permet pujar gran quantitat de dades amb força facilitat. Per l'usuari final és senzill accedir-hi i molt intuïtiu descarregar aquestes mateixes dades.

Tenint en compte que les dades que conformen els inventaris i catàlegs que l'ANC posa a disposició dels seus usuaris estan estructurades i, en un primer moment i abans de ser

migrades a GANC i a Arxius en Línia, estan en un format (access) relativament fàcil d'exportar a formats oberts, la plataforma CKAN podria ser un bon mitjà per, en un primer estadi, publicar dades obertes de l'Arxiu Nacional de Catalunya.

Així, a la pàgina web actual de l'ANC es podria informar a l'usuari de la possibilitat que té de descarregar-se en formats oberts les descripcions que l'Arxiu posa a la seva disposició a Arxius en Línia, oferint un enllaç que redirigís al mateix usuari al portal de dades obertes de l'ANC, desenvolupat amb CKAN.

Com a aspecte a millorar, i comprovant com queda la informació a CKAN amb els diferents formats i en determinats camps de cada registre, es podria plantejar la possibilitat d'oferir la informació en obert i continguda en les unitats de descripció dels inventaris i catàlegs d'una forma més estructurada. Així per exemple, s'ha pogut constatar com en el moment d'exportar a csv la informació que hi ha en els camps DESCRIPCIÓ o OBSERVACIONS de les unitats de catalogació de determinat fons o sèrie, en algunes d'elles la informació pot no quedar del tot ben representada, ja que en aquests camps hi ha salts de línia, barres i punts que poden distorsionar la representació del contingut que hi ha en aquests camps. Com s'ha dit anteriorment, en aquests camps hi consta concatenada informació que en la fitxa de primers tractaments de l'ANC està distribuïda en diferents camps. A l'hora d'oferir la informació en formats oberts es podria optar per no concatenar aquestes dades als camps esmentats i mantenir-la als camps originals de la fitxa de primers tractaments per tal que l'usuari final pogués treballar millor amb les dades. Això faria que, per exemple, l'usuari no hagués d'aïllar els topònims que poguessin constar al camp DESCRIPCIÓ ja que ja els tindria en un camp independent. Amb tot, es continuaria oferint a Arxius en Línia la informació distribuïda amb etiquetes en els camps DESCRIPCIÓ i OBSERVACIONS però al portal de dades obertes de l'Arxiu es podria donar la informació d'una manera més estructurada i aïllada.

La inversió, tant en recursos econòmics com humans, per dur a terme un projecte de dades obertes per l'Arxiu Nacional de Catalunya no seria molt elevada ja que el volum més gran de feina la realitzen els arxivers quan descriuen la documentació que es custodia a l'Arxiu o en el moment que es revisen i sistematitzen les dades contingudes en els instruments de descripció que els diferents arxius centrals departamentals de la Generalitat de Catalunya fan arribar a l'Àrea dels Fons de l'Administració de l'ANC. És evident que hi hauria d'haver una feina de disseny del portal de dades obertes de la institució, però una eina com CKAN facilitaria aquesta tasca.

En un futur, caldria plantejar-se la possibilitat de redissenyar Arxius en Línia, oferint la opció, des del mateix portal, d'exportar la informació a formats reutilitzables. Però fins que no

arribés aquest moment, l'ús de CKAN com a plataforma per oferir en formats reutilitzables les descripcions que l'ANC posa a disposició dels seus usuaris a la xarxa seria una molt bona alternativa.

5 Conclusions

Al llarg d'aquest treball s'ha fet un repàs a la legislació europea i estatal en relació a la reutilització de la informació del sector públic, s'ha analitzat el concepte *open data* i les possibilitats que ofereix i s'han observat diferents iniciatives *open* que estan duent a terme institucions arxivístiques d'arreu del món i com aquesta és una tendència creixent en el món dels arxius.

Amb l'anàlisi i el disseny del projecte d'*open data* per l'Arxiu Nacional de Catalunya i amb la posterior prova de concepte realitzada amb CKAN s'ha volgut demostrar que existeixen eines *open source* al mercat que permeten publicar dades estructurades en formats oberts d'una manera ràpida, senzilla i atractiva per les persones que hi estiguin interessades.

És indubtable que els arxius disposen de gran quantitat d'informació ben estructurada, certa, de qualitat i fàcilment exportable a formats reutilitzables. Com s'ha exposat en el treball, l'Arxiu Nacional de Catalunya podria oferir les mateixes dades que actualment posa a l'abast del ciutadà a Internet, en un altre tipus de format que facilités la tasca que realitzen els investigadors o els usuaris en general de l'Arxiu.

I ho podria fer:

- Bé a través d'una plataforma *open source* com CKAN: posant un enllaç a la pàgina web actual de l'Arxiu Nacional de Catalunya on s'informés a l'usuari que té a la seva disposició instruments de descripció en formats oberts en un altre recurs.
- Bé aportant aquestes dades al portal de dades obertes de la Generalitat de Catalunya: en aquest cas també es podria posar un enllaç a l'actual pàgina web de l'Arxiu que dugués a l'usuari a l'apartat del portal de dades obertes on hi hauria publicats els conjunts de dades de l'Arxiu Nacional de Catalunya.
- Bé actualitzant el portal Arxius en Línia: amb la col·laboració i la constitució de grups de treball amb arxivers de la resta d'arxius històrics de la Generalitat de Catalunya es podria analitzar com, des del portal Arxius en Línia, es podrien exportar els resultats obtinguts a partir de qualsevol cerca en formats reutilitzables.

L'obertura de dades tal i com s'ha plantejat en la prova de concepte feta sobre CKAN seria, potser, el pas més lògic que podria fer la institució en un primer moment. Més endavant, i en funció dels recursos humans i econòmics disponibles, es podria estudiar la possibilitat d'anar

més enllà, explorant les possibilitats d'utilitzar URI's per integrar les dades de l'ANC a la Web i enllaçar-les amb altres dades, dotant-les així de context. Analitzar, en definitiva, com utilitzar els documents, descripcions i dades de l'ANC en el marc de la web semàntica, per oferir els seus recursos, explotar-los i enriquir-los sota aquest estàndard.

6 Bibliografia i fonts

ALEIXANDRE-BENAVENT, R., FERRER.SAPENA, A., PESET, F. "Acceso a los datos públicos y su reutilización; open data y open government". *El profesional de la información*, vol. 20, núm. 3 (2011): 260-269. <http://eprints.rclis.org/20997/2/92741636q145x727.pdf>. (Consulta 30 de maig de 2016).

ARXIU NACIONAL DE CATALUNYA (Editorial). "El fons de l'Arxiu Nacional de Catalunya a internet". *Butlletí de l'Arxiu Nacional de Catalunya*, núm. 2 (2002): 1-3. (Consulta 8 de juny de 2016). <http://anc.gencat.cat/web/.content/anc/documents/arxiu/ANC02.pdf>.

ARXIU NACIONAL DE CATALUNYA (Editorial). "Estratègies de comunicació: el nou portal web de l'Arxiu Nacional de Catalunya. Una finestra al món". *Butlletí de l'Arxiu Nacional de Catalunya*, núm. 28 (2011): 1-2. (Consulta 8 de juny de 2016). <http://anc.gencat.cat/web/.content/anc/documents/arxiu/ANC28.pdf>.

BENETT, D. i HARVEY, A. "W3C Publishing Government Data". W3C Working Draft 2009. Consulta 30 de maig de 2016. <http://www.w3.org/TR/gov-data/>.

BERNAL CERCÓS, A. "Els arxius centrals han transferit més de 600.000 registres informàtics a l'ANC". *Butlletí Arxius*, núm. 52 (2008): 11. (Consulta 22 de juny de 2016). <http://continguts.cultura.gencat.cat/arxius/butlleti/hemeroteca/docs/arxius52.pdf>.

BERNAL CERCÓS, A. "El traspàs del SIGED al GANC. La descripció de l'arxiu de gestió arribar a l'arxiu històric". *Butlletí Arxius*, núm. 56 (2009): 3-6. (Consulta 22 de juny de 2016). <http://continguts.cultura.gencat.cat/arxius/butlleti/hemeroteca/docs/arxius56.pdf>.

BERNAL CERCÓS, A. "L'Arxiu Nacional de Catalunya". *Lligall. Revista Catalana d'Arxivística*, núm. 36 (2014): 69-76. (Consulta 6 de juny de 2016). <http://www.arxivers.com/index.php/documents/publicacions/revista-lligall-1/ligall-36-1/1340-04-els-arxius-del-sistema-darxius-de-catalunya/file>.

BERNERS-LEE, T., HENDLER, J., LASSILA, O. "The Semantic Web". *Scientific American*, vol. 284, núm. 5 (2001).

CARRILLO POZAS, A. "Retos y oportunidades en Archivos y Gestión documental ante la Web Semántica". Curs impartit a l'Escola Gallega d'Administració Pública, 21-22 abril, 2014.

Consulta 21 de juny de 2016. <http://es.slideshare.net/bne/retos-y-oportunidades-en-archivos-y-gestin-documental-ante-la-web-semantica-34386038>.

CASTAN, A.; PETIT, C.; UTGÉS, M. "El nou cercador web de l'Arxiu Nacional de Catalunya". *Butlletí Arxius*, núm. 57 (2009): 9-13. (Consulta 22 de juny de 2016). <http://continguts.cultura.gencat.cat/arxius/butlleti/hemeroteca/docs/arxius57.pdf>.

CHIARETTI, A. "SAN-LOAD. El Sistema Archivístico Nacional Italiano como Datos Abiertos Enlazados". Ponència presentada al Congrés Internacional Innova Tics, Santiago de Chile, 27-28 agost, 2014. Consulta 21 de juny de 2016. <http://congresoinnovatics.org/content/alessandro-chiaretti-istituto-centrale-gli-archivi#overlay=node/91/edit>.

DIETRICH, D., GRAY, J., MCNAMARA, T., POIKOLA, A., POLLOCK, R., TAIT, J. i ZIJLSTRA, T. "The Open Data Handbook". Consulta 30 de maig de 2016. <http://opendatahandbook.org/guide/en/>.

DE LOS RÍOS, N. "Gobierno Abierto". *Lligall. Revista Catalana d'Arxivística*, núm. 35 (2013): 139-150. (Consulta 6 de juny de 2016). <http://www.arxivers.com/index.php/documents/publicacions/revista-lligall-1/lligall-35-1/1354-08-gobierno-abierto/file>.

FERRER, J., MAGRINYÀ, A., SERNA, E. "Arxius en línia: una realitat en transició cap el Portal d'Arxius Catalans". Ponència presentada al Congrés d'Arxius i Indústries Culturals, Girona, 11-15 octubre, 2014. Consulta 9 de juny de 2016. <http://www.girona.cat/web/ica2014/ponents/textos/id211.pdf>.

FUNDACIÓN ORANGE. "Datos abiertos en la Comunidades Autónomas y sus mayores ayuntamientos 2014". (2014). Consulta 13 de juny de 2016. https://www.proyectosfundacionorange.es/docs/eE2014/Datos_Abiertos_2014.pdf.

GÓMEZ DUEÑAS, L.F. "La Iniciativa de Archivos Abiertos (OAI), un nuevo paradigma en la comunicación científica y el intercambio de información". *Revista Códice*, núm. 5 (2005): 4-25.

MALAMUD, C. et al. "Open Government Data Principles". 2007. Consulta 25 de maig de 2016. https://public.resource.org/8_principles.html.

MATAS i BALAGUER, J. "Notes sobre les lleis de transparència". *Lligall. Revista Catalana d'Arxivística*, núm. 38 (2016): 22-44. (Consulta 6 de juny de 2016).

PETIT CIBIRAIN, C. "El lloc web de l'Arxiu Nacional de Catalunya". *Butlletí de l'Arxiu Nacional de Catalunya*, núm. 4 (2003): 7. (Consulta 8 de juny de 2016). <http://anc.gencat.cat/web/.content/anc/documents/arxiu/ANC04.pdf>.

SÁNCHEZ MAIRENA, A. "Acercando los archivos a los ciudadanos. Una experiencia desde el Portal de Archivos Españoles (PARES) del Ministerio de Educación, Cultura y Deporte". *Arbor*, vol. 190, núm. 765 (2014). (Consulta 21 de juny de 2016). <http://arbor.revistas.csic.es/index.php/arbor/article/viewArticle/1905/2131#NOTE01>.

WONDERLICH, J. "Ten Principles for Opening Up Government Information". 2010. Consulta 25 de maig de 2016. <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>.